A REVIEW ON PLANT RECOGNITION AND CLASSIFICATION TECHNIQUES USING LEAF IMAGES

Anant Bhardwaj[*1], Manpreet Kaur[2]

[1]*M.Tech Scholar, Department of Electronics and Communication Engineering, Lovely Professional University INDIA*
[2]*Assistant Professor, Department of Electronics and Communication Engineering, Lovely Professional University INDIA*

[*]*Corresponding Author: Anant Bhardwaj*

Abstract- **Automatic digital plant classification and retrieval can be achieved by extracting features from its leaves. There are various opportunities to improve plant species identification due to computerization through the designing of a convenient automatic plant recognition system. Many different approaches consist of some major parts. First, images of leaf are acquired with digital camera or scanners. Then the user can selects the base point of the leaf and a few reference points on the leaf blades or done this automatically. Then several morphological features are extracted. These features are used as inputs to the classifier system for discrimination as probabilistic neural network. The network was trained with leaves from different plant species. Then the recognition accuracy of the proposed method has been tested. The method works only for the plants with broad flat leaves which are more or less two dimensional in nature. This paper presented various effective algorithms used for plant classification using leaf images and review the main computational, morphological and image processing methods that have been used in recent years and we conclude with a discussion of ongoing work and outstanding problems in the area.**

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

I. INTRODUCTION

Plant is one of the most important forms of life on earth. Plants maintain the balance of oxygen and carbon dioxide of earth's atmosphere. The relations between plants and human beings are also very close. In addition, plants are important means of livelihood and production of human beings. Unfortunately, the overwhelming development of human civilization has disrupted this balance to a greater extent than we realize. It is one of our biggest responsibilities to save the plants from various threats, restore the diverseness of the plant community and put everything back to balance. The first step of protecting plants is to automatically recognize or classify them means understand what they are and where they come from. But it is very difficult for ones to recognize a plant in hand correctly and immediately because there are so many kinds of plants unknown to us on earth. Allopathic medicine has a rich history with a number of ayurvedic [15] leaves which can't be recognized by a human being. A computerized plant identification system can be very helpful in botanical garden or natural reserve park management, new plant species discovery, plant taxonomy, exotic plant detection, edible/poisonous plant identification and so on.

A computer based plant identification or classification system can use different characteristics of the flora, starting at very simple level such as: shape and colour of the leaf, flower and fruit type, branching style, root type, seasonality, outlook, to very complex such as cell and tissue structure, genetic structure. Presently the cell phones are capable of acquiring high quality images with their integrated digital camera, which makes the usability of such a system even wider. Plant leaves are two dimensional in nature and hold important features that can be useful for classification of various plant species. There are set of appropriate numerical attributes of features to be extracted from the object of interest for the purpose of classification. Research on the utilization of moments for object characterization in both in- variant and non-invariant tasks has received considerable attention in recent years [8, 9]. The mathematical concept of moments has been around for many years and has been used in many diverse fields ranging from mechanics and statistics to pattern recognition and image understanding.

In general, it is not practical for a normal human being to examine and recognize the large amount of plant species and moreover it is extremely cost effective. Hence the system is required to overcome the errors caused by conventional plant species recognition systems which based solely on human expertise. Requirement of better plant species recognition system is in great demand and several researches have been made by some authors which are discussed in the next section.

II. TECHNIQUES USED IN PLANT RECOGNITION

A substantial amount of work has been done on leaf shape based plant classification and recognition. However, there are still many approaches could be investigated to develop robust plant identification systems. A certain method may give good performance in specific samples of leaves, but does not guarantee to perform good result for other ones. Sometime it is found that two or more plants have leaves with similar or same shape, but different colors. In that case, color features cannot be neglected. A typical image based plant identification system is shown in Fig. 1.

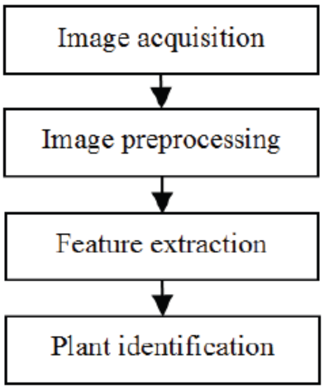

Figure 1. System Flowchart

Wang et. al. [2] employed centroid contour distance (CCD) curve, eccentricity and angle code histogram (ACH). Their experimental results on 1400 leaf images from 140 plants show that the proposed approach can achieve a better retrieval performance than both the curvature scale space (CSS) method and the modified Fourier descriptor (MFD) method. Fu et al. [3] also used centroid-contour distance curve to represent leaf shapes in which an integrated approach for an ontology-based leaf classification system is proposed, where in machine learning techniques play a crucial role for the automatization of the system. For the leaf contour classification, a scaled CCD code system is proposed to categorize the basic shape and margin type of a leaf by using the similar taxonomy principle adopted by the botanists. Then a trained neural network is employed to recognize the detailed tooth patterns. For the leaf vein recognition, the vein texture is extracted by employing an efficient combined thresholding and neural network approach so as to obtain more vein details of a leaf.

Plant leaves are approximately 2D in nature and the shape of plant leaf is one of the most important features for characterizing various plants species. Therefore, it is necessary for us to develop an automatic easy method that can correctly recognize leaf shapes of different species. Du et al. [4] an efficient computer-aided plant species identification (CAPSI) approach is proposed, which is based on plant leaf images using a shape matching technique. Firstly, a Douglas-Peucker approximation algorithm is adopte to the original leaf shapes and a new shape representation is used to form the sequence of invariant attributes. This is shape polygon approximation. Then a modified dynamic programming (MDP) algorithm for shape matching is proposed for the plant leaf recognition. Finally, the superiority of proposed method is demonstrated by experiment. In this experiment, there are 50 leaf images randomly selected from our image database as the query images, and each query can retrieve the 20 most similar images from the database. For recognition of intact leaves, the recognition accuracy of the MDP is over 92%.

Gu et al. [5] used the result of segmentation of leaf's skeleton based on the combination of wavelet transform (WT) and Gaussian interpolation. It is a a new approach for leaf recognition also using the classifiers, a nearest neighbour classifier (1-NN), a k -nearest neighbor classifier (k-NN) and a radial basis probabilistic neural network (RBPNN) are used, based on run-length features (RLF) extracted from the skeleton to recognize the leaves. Finally, the effectiveness and efficiency of the proposed method is demonstrated by several experiments.

Wang et al. [6] extracted several geometric features like rectangularity, circularity, eccentricity and seven moment invariants for classification. Hu defines the following seven functions, computed form central moments through order three, that are invariant with respect to object scale, translation and rotation:

$$\phi_1 = \mu_{20} + \mu_{02} \quad (1)$$

$$\phi_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^{\ 2} \quad (2)$$

$$\phi_3 = (\mu_{30} - 3\mu_{12})^2 + 3(\mu_{21} + \mu_{03})^2 \quad (3)$$

$$\phi_4 = (\mu_{30} - \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \quad (4)$$

$$\phi_5 = (\mu_{30} + 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \quad (5)$$

$$\phi_6 = (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11})(\mu_{30} + \mu_{12}) - (\mu_{21} + \mu_{03}) \quad (6)$$

$$\phi_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \quad (7)$$

15 features are extracted from pre-processed leaf. In that paper he proposed an iterative threshold selection segmentation method to address leaf images with simple background and Marker-controlled watershed segmentation method is selected for those leaf images with complicated background. In addition, a moving center hypersphere (MCH) classifier based on shape features is proposed for classifying a large number of leaves. If we use conventional classifier like K-NN or Neural Network, the corresponding classification process would be quite time-consuming and space-consuming. We regard each class of patterns as a series of "hyper-spheres", while in conventional approaches these patterns from one class are all

treated as a set of “points”. As a result there are more than 20 classes of plant leaves successfully classified. The average correct recognition rate is up to 92.2 percent.

Du et al. [7] introduced shape recognition based on radial basis probabilistic neural network (RBPNN) which is trained by orthogonal least square algorithm (OLSA) and optimized by recursive OLSA. It performs plant recognition through modified Fourier descriptors of leaf shape. It can identify the type of plant from a partially damaged or broken leaf. A set of unique features with proper proposed method is responsible for plant identification. The experimental result shows that the RBPNN achieves higher recognition rate and efficiency than radial basis function neural network (RBFNN), BP neural network (BPNN).

Historically, Hu[8] published the first significant paper on the utilization of moment invariants for image analysis and object representation in 1961. Hu’s Uniqueness Theorem states that if f(x, y) is piecewise continuous and has nonzero values only 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 $\{m_{pq}\}$ is uniquely determined by f(x, y) and conversely f(x,y) is uniquely determined by $\{m_{pq}\}$. Since an image segment has finite area and, in the worst case, is piecewise continuous, a moment set can be computed and used to uniquely describe the information contained in the image segment. Using nonlinear combinations of geometric moments, Hu derived a set of invariant moments which has the desirable properties of being invariant under image translation, scaling, and rotation [1]. However, the reconstruction of the image from these moments is deemed to be quite difficult. T.H.Resis [9] stated that moment invariant for pattern recognition presented by HU is incorrect. The four moment absolute invariant under general linear transformation is in error. So he revised the fundamental theorem. Sidhartha Maître [10] consider the change of effect of contrast in an image and modified the seven moment invariant given by HU which are independent of change of transformation, scale, rotation and contrast also. Similarly Jan Fusser [11] proposed a new set of moment invariants with respect to rotation, translation, and scaling suitable for recognition of objects having -fold rotation symmetry.

Wu et al. [12] extracted 12 commonly used digital morphological features which were orthogonalized into 5 principal variables. They used 1800 leaves to classify 32 kinds of plants using probabilistic neural network (PNN) system. The major contribution of this work is a easy to implement method. First, images of leaf are acquired with digital camera or scanners. Then the system requires the user to select the base point of the leaf and a few reference points on the leaf blades, in which first we convert the RGB image [1] to binary image using the given formula.

$$gray = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (8)$$

where R, G, B correspond to the colour of the pixel, respectively.

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

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

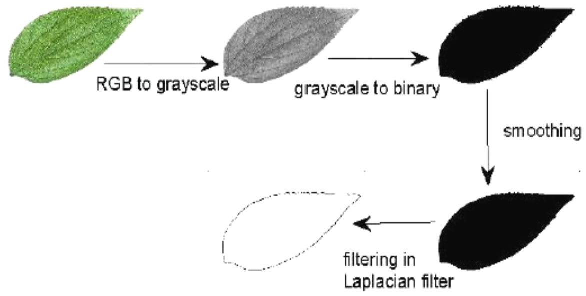

Figure 2. Image pre-processing steps

After pre-processing is complete as shown in Fig.2, obtain a set of basic geometric features such as Diameter, Physiological length, Physiological width, leaf area and leaf perimeter and after that 12 morphological features as smooth factor, aspect ratio, form factor, rectangularity, narrow factor, perimeter ratio of diameter, perimeter ratio of Physiological length, Physiological width and 5 vien features were extracted from the leaves. After that transform the data to a new coordinate using PCA [19] i.e. use the mapping of $R^{12} \rightarrow R^5$. The PNN [18] as shown in Fig.3 is trained by 1800 leaves to classify 32 kinds of plants with good accuracy.

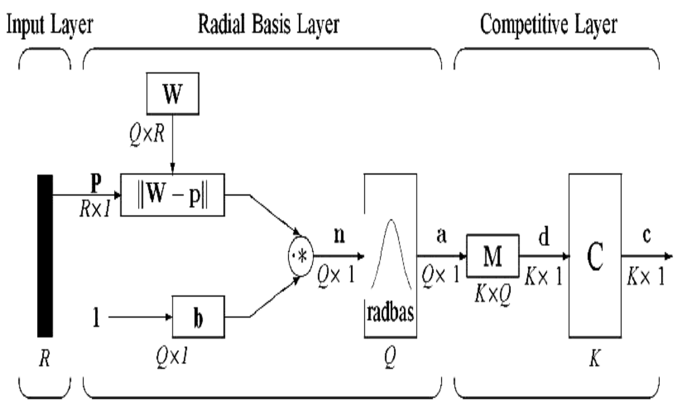

Figure 3. Network Structure, R=5, Q= 1800, K=32

Du and Zhang [13] approach to a new classification method, named as move median centers (MMC) hypersphere classifier, for the leaf database based on digital morphological feature is proposed. They well explained the Algorithm of MMC, classification stage and the Data pre-processing for application of the MMC. In particular, by comparing with the nearest neighbour (1-NN) and k-NN classifiers, it can be found that the MMC classifier can not only save the storage space but also reduce the classification time. The proposed method is more robust than the one based on contour features since those significant curvature points are hard to find.

An uncomplicated and computationally effective technique for plant species recognition by means of leaf image is recommended by Hossain and Amin [14]. This technique executes only on the plants with wide flat leaves which are more or less two dimensional in general. This technique includes five major phases. In first phase, images of leaf are obtained with digital camera or scanners. Subsequently the user decides the base point of the leaf and small number of reference points on the leaf blades. In accordance with these points the leaf shape is obtained from the background and a binary image is generated. Following that the leaf is aligned horizontally by keeping its base point on the left of the image. Then eccentricity, area, perimeter, major axis, minor axis, equivalent diameter, convex area and extent are obtained as before done. A distinctive collection of features are obtained from the leaves by segmenting across the major axis and parallel to the minor axis. After that the feature points are normalized by considering the ratio of the slice lengths and leaf lengths (major axis). These features are provided as inputs to the probabilistic neural network [18].

A new technique for feature extraction from natural image like plant leaf is developed by Prasad et al. [16] for automated living plant species identification 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, precisely in order that the orientation of the object in the image does not taken into account and which also enhance the accuracy rate. These coefficients are given as the input to a trained SVM classifier to categorize the result.

Abdul kadir [17] build a foliage plant identification systems for 60 kinds of leaves . It was dedicated to handle two or more plants that have similar/same shape but the color patterns on the leaves were different. In this case, Zernike moments were combined with other features: geometric features, color moments and gray-level co-occurrence matrix (GLCM). To implement the systems, two approaches have been investigated. First approach used a distance measure and the second used Probabilistic Neural Networks (PNN). The results show that Zernike Moments have a prospect as features in leaf identification systems when they are combined with other features.

Table 1. An Overview of the various plant classification techniques

S.NO.	RESEARCHERS	METHODS / TECHNIQUES USED	REFERENCE
1	Z. Wang, Z. Chi, and D. Feng	Curvature scale space (CSS) and modified Fourier descriptor (MFD) method.	[2]
2	H. Fu, Z. Chi, D. Feng, and J. Song	Machine learning techniques for ontology-based leaf classification and centroid-contour distance curve method	[3]
3	J. Du, D. Huang, X. Wang, and X. Gu	Computer-aided plant species identification (CAPSI) approach with Douglas-Peucker approximation and modified dynamic programming (MDP) algorithm	[4]
4	X. Gu, J. Du, and X. Wang,	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)	[5]
5	X. Wang, J. Du, and G. Zhang	Geometric features, 7 moment invariants for classification and hypersphere classifier	[6]
6	J. Du, D. Huang, X. Wang, and X. Gu	Radial basis probabilistic neural network which is trained by orthogonal least square algorithm (OLSA) and optimized by recursive OLSA.	[7]
7	S. Wu, F. Bao, E. Xu, Y. Wang, Y. Chang, and Q. Xiang	PCA technique and training, testing with Probabilistic Neural Network	[12]

8	J.-X. Du, X.-F. Wang and G.-J. Zhang	Move median centers (MMC) hypersphere classifier	[13]
9	Hossain and M.A. Amin	Features are selected after 5 major phses and used as inputs to the probabilistic neural network	[14]
10	S. Prasad, P. Kumar and R.C. Tripathi	New multi-resolution and multidirectional curvelet transform is applied on subdivided leaf images to extract leaf information	[16]
11	Abdul Kadir, L.E. Nugroho, A.susanto, P.Insap Santosa	Zernike moments with other features: geometric features, colour moments and gray-level co-occurrence matrix (GLCM)	[17]

III. PROBLEMS AND DIRECTIONS

There are several issues in the above discussed plant species recognition systems in providing better classification results since the number of varieties of plants and its features are large in number. In order to overcome these difficulties numerous researches has to be carried out in future. The research may be focussed on the following areas.

- Together with the macroscopic features of plants, physical features like weight, texture, size, colour, odour and hardness can be taken into account for better classification results.
- In order to implement a consistent plant species recognition system, a representative more and more database must be assembled in order that machine learning techniques can be trained accurately.
- Better plant species recognition system can be designed like automatic visual inspection systems by integrating techniques like image processing, statistical feature extraction and machine learning techniques.
- On the other hand conventional machine learning datasets typically have only a lesser number of handful classes. In addition, the conventional classifiers fundamentally return the best matching class, even if the possibility of a match is extremely low.

IV. CONCLUSION

In this paper we have discussed a number of plants identification systems with wide ranges of methods. It should be clear that no single method provides a solution for all problems, but appropriate methods must be chosen. There are many ways to recognize the plant species, but it usually time consuming. This is because it involves the expert, botanists to recognize the plant species. Therefore, it is very important to automatically recognize the plant species in order to manage them. This will increase an interest in studying plant taxonomy and ecology, lift biology educations standards and promote the use of information technology for the management of natural reserve parks and forest plantation. By doing this, the best approach for leaf features extraction and classification have to be analyze.

Plant species recognition system is commonly used in agriculture, ecology and environmental science. Comparing to other recognition system, plant species recognition system requires additional skilled understanding. Accordingly, the requirement of plant species recognition system using computer vision techniques is increasing rapidly for several applications. Although performance of the system is good enough, we believe that the performance still can be improved. Hence, other features will be researched in the future.

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 (ICCIT), 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 (IJCSE), 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 (ICCCT), 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] V. Cheung and K. Cannons, "An Introduction to Probabilistic Neural Networks", http://www.psi.toronto.edu/~vincent/research/presentations/PNN.pdf, (2003).

[19] J. Slens, "A Tutorial on Principal Component Analysis", http://www.snl.salk.edu/~shlens/pca.pdf, (2009).