International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 6, Volume 2 (June 2015) www.ijirae.com

A Review of Plant Recognition Methods and Algorithms

Sapna Sharma[#1], Dr. Chitvan Gupta[#2]
CSE Department, Noida Institute of Engineering and Technology, Gr.Noida, Uttar Pradesh, INDIA-201306

Abstract- Plants are the one of the major source of medicine. Now days this is hard to recognise a plant by looking its physical properties. A specialist of this filed may recognise the plant but it is very difficult to a lay man to recognise a plant of rare category. Artificial intelligence is the one the major research area which uses leaf features such as leaf contour, eccentricity, centroid, solidity, etc as an input to the any system and classified any plant accurately. Using different software tool this process can be automated. Various techniques such as principle component analysis (PCA), Hu's moment invariant method and morphological features based analysis are also used by various researchers to provide an efficient and highly accurate automated tool for recognition of plant using leaf images.

Keywords: Plant recognition Neural Network, Artificial Intelligence, Leaf image, feature extraction, PCA, morphological features, Moment invariant.

I. INTRODUCTION

Plants are an integral part of all natural life [1] and systematically classifying those helps ensure the protection and survival of all natural life. Plants are also important for their medicinal properties, as alternative energy sources like bio-fuel and for meeting our various domestic requirements like timber, clothing, food and cosmetics. A computerized plant identification system can be very helpful in botanical garden or natural reserve park management, 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.

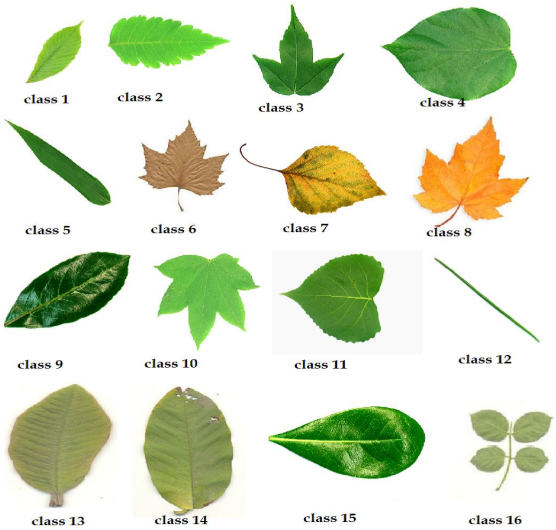
Classifying plants using leaves usually uses descriptions by botanists who describe the different features including shape, texture and veins [1], [2], [3], [4]. Plant identification is generally based on the observation of the morphological characteristics of the plant (such as general character, structures of stems, roots and leaves, embryology and flowers) followed by the consultation of a guide or a known database. An important amount of information about the taxonomic identity of a plant is contained in its leaves. Moreover, leaves are present on the plants for several months in a year, whereas flowers and fruits may remain only several weeks. This is why most plant identification tools based on Content-Based Image Retrieval techniques work on leaf image databases [5][6]. A leaf can be characterized by its color, its texture, and its shape. The color of a leaf may vary with the seasons and climatic conditions. A first group of methods extracts morphological plant characters commonly used in botany. Du et al. [7] compute eight features, Aspect Ratio, Rectangularity, Convex Area Ratio, Convex Perimeter Ratio, Sphericity, Circularity, Eccentricity and Form Factor, from the boundary of the leaves. In [8] Hu's established a fundamental theorem to relate two-dimensional moment invariant to well known algebraic invariants. Here Hu shows that recognition of geometrical patterns and alphabetical characters independently of position, size and orientation can be accomplished. In [11], Sandeep Kumar has paper presents an approach where the plant is identified based on its leaf features such as area, color histogram edge histogram and the Gray-Level Co-occurrence matrix (GLCM) and Principal Component Analysis (PCA) algorithms have been considered to extract the leaves features in new plant.

TABLE I: AN OVERVIEW OF SOME EXISTING METHOD OF PLANT LEAF BASED RECOGNITION SYSTEMS

S.NO.	RESEARCHERS	METHOD USED IN PLANT RECOGNITION
1	Sandeep Kumar.E	Proposed a leaf color, area and edge features based approach for identification of Indian medicinal plants.[11]
2	Abdolvahab Ehsanirad and Sharath Kumar Y. H.	Proposed GLCM and PCA methods [12]
3	Ming-Kuei Hu	Proposed moment invariant (MI) based approach and results show that recognition schemes based on these invariants could be truly position, size and orientation independent, and also flexible enough to learn almost any set of patterns. [8]
4	Xiao Gu , Ji-Xiang Du , and Xiao-Feng Wang	Leaf Recognition Based on the Combination of Wavelet Transform and Gaussian Interpolation1

5	Hong Fu and Zheni Chi	Two-stage approach for leaf vein extraction. Combined a thresholding method and an Artificial Neural Network (ANN) classifier to extract leaf veins[13]
6	Zalikha Zulkifli et al.,	Features are extracted from the effective moment invariant technique and classified using the General Regression Neural Network (GRNN) [14]
7	Prasad et al.,	New multi-resolution and multidirectional curvelet transform is applied on subdivided leaf images to extract leaf information [15]
8	Zheru Chi et al.,	Novel style of Gabor filter banks designed for plant species recognition using their bark texture features [16]
9	Yusof et al.,	Used Gabor filter in the pre-processing stage of the wood texture image to multiply the number of features for a single image, thus providing more information for feature extractor to capture [17]
10	Marzuki Khalid et al.,	Wood features are extracted based on two feature extractors: Basic Grey Level Aura Matrix (BGLAM) technique and statisticalproperties of pores distribution (SPPD) technique [18]
11	Runtz et al.	Developed a real-time plant recognition algorithms and associated electronic hardware [19]
12	Lei Zhang et al.	Self-Organizing feature Map (SOM) neural network to identify the plant species [20]
13	A.H. Kulkarni Dr. H.M.Rai Dr. K.A.Jahagirdar P.S.Upparaman	Suggest Radial Basis Probabilistic Neural Network (RBPNN) and Zernike moments for plant classification [21]
14	Javed Hossain, M. Ashraful Amin	Leaf Shape Identification Based Plant Biometrics in which extracted parameters of leaf and a typical probabilistic neural network is used [22]
15	Jyotismita Chaki and Ranjan Parekh	Proposed a binary superposition (B-S) approach and also compare this approach to MI and CR based approach [23]

II. CLASSIFICATION OF PLANTS LEAF



III. DIFFERENT APPROCHES OF FEATURE EXTRACTION

This approach uses 11 digital and morphological features which are briefly explained as:

1. **Major axis length:** The line segment connecting the base and the tip of the leaf is the major axis
2. **Minor axis length:** The maximum width, which is perpendicular to the major axis, is the minor axis of a leaf.
3. **Convex area:** It specifies the number of pixels in ‘Convex Image’.
4. **Filled area:** The total number of on pixels in Filled Image is known as Filled Area.

5. **Eccentricity:** The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1. The eccentricity is defined as Eccentricity = w/l where w is the length of the minor axis and l is the length of the leaf' major axis. This feature can be used to differentiate the rounded leaf and the long one.
6. **Perimeter:** The distance around the boundary of the region is called perimeter. The command 'regionprops' in MATLAB computes the perimeter by calculating the distance between each adjoining pair of pixels around the border of the region.
7. **Solidity:** The proportion of the pixels in the convex hull that are also in the region. It is computed as Area/Convex Area.
8. **Orientation:** The angle (in degrees ranging from -90 to 90 degrees) between the x-axis and the major axis of the ellipse that has the same second-moments as the region.
9. **Extent:** Extent specifies the ratio of pixels in the region to pixels in the total bounding box. It is computed as the Area divided by the area of the bounding box.
10. **Centroid:** It specifies the center of mass of the region. The first element of Centroid is the horizontal coordinate (or x-coordinate) of the center of mass & the second element is the vertical coordinate (or y-coordinate).
11. **Equiv Diameter:** It is the diameter of a circle with the same area as the region. It is Computed as sqrt (4*Area/pi).

SOME OTHER FEATURES WHICH CAN ALSO BE EXTRACTED FROM LEAF IMAGE ARE AS FOLLOWS:

12. **Roundness** The roundness or circularity ratio is defined as Roundness=A/P^2, where A is the area of the leaf and P is the perimeter of the leaf. This feature also can be used to differentiate the rounded leaf and the long one.
12. **Dispersion:** Dispersion is ratio between the radius of the maximum circle enclosing the region and the minimum circle that can be contained in the region. Mathematically, it is notated as below

$$dispersion = \frac{max(\sqrt{(x_i - x')^2 + (y_i - x')^2})}{min(\sqrt{(x_i - x')^2 + (y_i - x')^2})}$$

In the above formula, (x',y') is the centroid of the leaf and (x_i, y_i) is the coordinate of a pixel in the leaf contour.
13. **Color Moments:** Color moments represents color features that are extracted from color information on the leaf by using statistical calculations such as mean (μ), standard deviation (σ), skewness (θ), and kurtosis (γ). The four features are calculated as follows:

$$\mu = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(P_{ij})$$

$$\sigma = \sqrt{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(P_{ij-}\mu)^2}$$

$$\theta = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(P_{ij-}\mu)^3}{mn\sigma^3}$$

$$\gamma = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(P_{ij-}\mu)^4}{mn\sigma^4}$$

Where, m is the height of the image, n is the width of the image, and Pij is the value of colour on rows i and column j.

14. **Texture Features:** This research used GLCM to capture textural information in the leaf. Five features were derived from GLCM:
    - Angular Second Moment (ASM),
    - Contrast
    - Inverse Different Moment (IDM)
    - Entropy
    - Correlation.

IV. CLASSIFIER ALGORITHMS

1. **Principal Component Analysis (PCA):** PCA is used to reduce the dimension of input vector of neural network. The purpose of PCA is to present the information of original data as the linear combination of certain linear irrelevant variables. Mathematically, PCA transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate, the second greatest variance on the second coordinate, and so on. Each coordinate is called a principal component[9].

The objective of PCA is to perform dimensionality reduction while preserving as much of the randomness in the high-dimensional space as possible But the limitation with PCA is it depends on scaling of variables and it is not always easy to interpret principal components. The main limitation of PCA is that it does not consider class separability since it does not take into account the class label of the feature vector.[10]

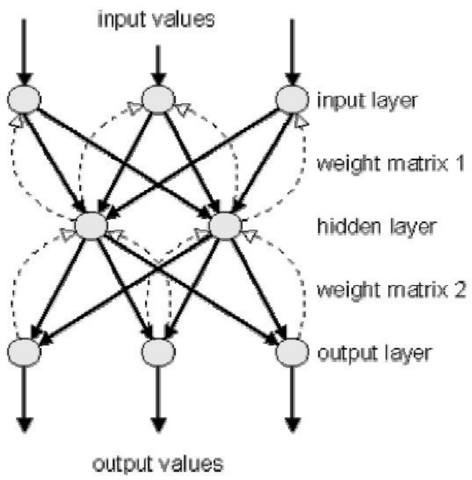
2. Moment Invariant Method (MI):

Moments and functions of moments have been extensively employed as invariant global features of images in pattern recognition[24]. Image moment is a certain particular weighted average (moment) of the image pixel' intensities, or a function of such moments, usually chosen to have some attractive property or interpretation. The idea of using moments in shape recognition gained prominence when Hu, derived a set of seven invariants using algebraic invariants [25].

3. Probabilistic Neural Network (PNN):

A probabilistic neural network (PNN) is a feed forward, which was derived from the Bayesian network and a statistical algorithm called Kernel Fisher discriminant analysis. It was introduced by D.F. Specht in the early 1990s.[3] In a PNN, the operations are organized into a multilayered feed forward network with four layers:

- Input layer
- Hidden layer
- Pattern layer/Summation layer
- Output layer



PNN is often used in classification problems. When an input is present, the first layer computes the distance from the input vector to the training input vectors. This produces a vector where its elements indicate how close the input is to the training input. The second layer sums the contribution for each class of inputs and produces its net output as a vector of probabilities. Finally, a compete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 (positive identification) for that class and a 0 (negative identification) for non-targeted classes.

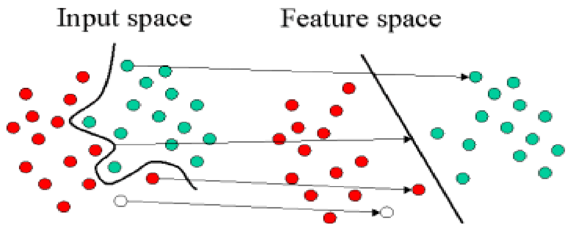
SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is primarily a classier method that performs classification tasks by constructing hyper planes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables. For categorical variables a dummy variable is created with case values as either 0 or 1. Thus, a categorical dependent variable consisting of three levels, say (A, B, C), is represented by a set of three dummy variables:

A: {1 0 0}, B: {0 1 0}, C: {0 0 1}

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. Support vector machine (SVM) is a non-linear classifier. The idea behind the method is to non-linearly map the input data to some high dimensional space, where the data can be linearly separated, thus providing great classification performance. SuThe illustration below shows the basic idea behind Support Vector Machines. Here we see the original objects (left side of the schematic) mapped, i.e., rearranged, using a set of mathematical functions, known as kernels.

The process of rearranging the objects is known as mapping (transformation). Note that in this new setting, the mapped objects (right side of the schematic) is linearly separable and, thus, instead of constructing the complex curve (left schematic), all we have to do is to find an optimal line that can separate the GREEN and the RED objects.



Genetic Algoruthms(GA) :Genetic Algorithms are mainly used for feature classification and feature selection. The basic purpose of genetic algorithms (GAs) is optimization. A genetic algorithm selects high strength classifiers as "parents," forming "offspring" by recombining components from the parent classifiers. The offspring displace weak classifiers in the system and enter into competition, being activated and tested when their conditions are satisfied. Thus, a genetic algorithm crudely, but at high speed, mimics the genetic processes underlying evolution. The basic purpose of genetic algorithms (GAs) is optimization. Since optimization problems arise frequently, this makes GAs quite useful for a great variety of tasks. As in all optimization problem, we are faced with the problem of maximizing/minimizing an objective function f(x) over a given space X of arbitrary dimension. A brute force which would consist in examining every possible x in X in order to determine the element for which f is optimal is clearly infeasible. GAs give a heuristic way of searching the input space for optimal x that approximates brute force without enumerating all the elements and therefore bypasses performance issues specific to exhaustive search.

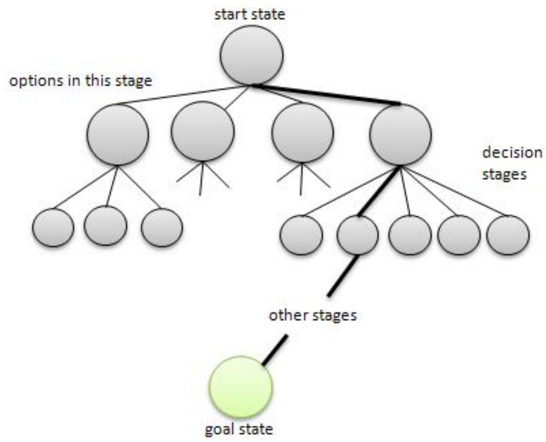
k-NEAREST NEIGHBOUR CLASSIFIERS

Knearestneighbor (kNN) classification is one of the most fundamental and simple classification methods and should be one of the first choices for a classification study when there is little or no prior knowledge about the distribution of the data. K-nearest-neighbor classification was developed from the need to perform discriminant analysis when reliable parametric estimates of probability densities are unknown or difficult to determine Nearest Neighbor classifier calculates the minimum distance of a given point with other points to determine its class.

Let us take an example. According to the k-NN rule suppose we first select k = 5 neighbors of w. Because three of these five neighbors belong to class 2 and two of them to class 3, the object w should belong to class 2, according to the k-NN rule. It is intuitive that the k-NN rule doesn't take the fact that different neighbors may give different evidences into consideration. Actually, it is reasonable to assume that objects which are close together (according to some appropriate metric) will belong to the same category. According to the k-NN rule suppose we first select k = 5 neighbors of w. Because three of these five neighbors belong to class 2 and two of them to class 3, the object w should belong to class 2, according to the k-NN rule.

Decision Tree

A decision tree is a classifier expressed as a recursive partition of the instance space. The decision tree consists of nodes that form a rooted tree, meaning it is a directed tree with a node called “root” that has no incoming edges. All other nodes have exactly one incoming edge. A node with outgoing edges is called an internal or test node. All other nodes are called leaves (also known as terminal or decision nodes). In a decision tree, each internal node splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes values. In the simplest and most fre-quent case, each test considers a single attribute, such that the instance space is partitioned according to the attribute’s value. In the case of numeric attributes, the condition refers to a range. Each leaf is assigned to one class representing the most appropriate target value. Alternatively, the leaf may hold a probability vector indicating the probability of the target attribute having a certain value. Instances are classified by navigating them from the root of the tree down to a leaf, according to the outcome of the tests along the path. Figure 9.1 describes a decision tree that reasons whether or not a potential customer will respond to a direct mailing. Internal nodes are represented as circles, whereas leaves are denoted as triangles. Note that this decision tree incorporates both nominal and numeric attributes. Given this classifier, the analyst can predict the response of a potential customer (by sorting it down the tree), and understand the behavioral characteristics of the entire potential customers population regarding direct mailing. Each node is labeled with the attribute it tests, and its branches are labeled with its corresponding values.



V. CONCLUSION

Plant species recognition system is extensively used in agriculture, ecology and environmental science. Comparing to other recognition system, plant species recognition system requires additional skilled understanding. It means that without any professional knowledge in this field it is almost impractical for others to classify a plant category in the level of species. Accordingly, the requirement of plant species recognition system using computer vision techniques is increasing rapidly for several applications. The accurate recognition of plant species is of essential requirement for the agriculture based industries and it has numerous additional applications. As discussed in the previous section, machine learning techniques can be used to accurately classify the plants.

VI. REFERENCES

[1]. S. G. Wu, F. S. Bao, E. Y. Xu, Y. Wang, Y.-F. Chang, and Q.-L. Xiang,“A leaf recognition algorithm for plant classification using probabilistic neural network,” CoRR, vol. abs/0707.4289, 2007.
[2]. Z. Wang, Z. Chi, and D. Feng, “Shape basedleaf image retrieval,” IEEE Proceedings-Vision, Image and Signal Processing, vol. 150,no. 1, February 2003.
[3]. X. Wang, D. Huang, J. Dua, H. Xu, L. Heutte, “Classification of plant leaf images with complicated background”, Special Issue on Advanced Intelligent Computing Theory and Methodology in Applied
[4]. Mathematics and Computation, Volume 205, Issue 2, Pages 916-926, 15 November 2008.
[5]. 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.
[6]. P. Belhumeur, D. Chen, S. Feiner, D. Jacobs, W. Kress, H. Ling, I. Lopez, R. Ramamoorthi, S. Sheorey, S. White, and L. Zhang. Searching the world’s herbaria: A system for visual identification of plant species. In European Conference on ComputerVision (ECCV), pages 116–129, 2008.
[7]. O. M. Bruno, R. de Oliveira Plotze, M. Falvo, and M. de Castro. Fractal dimension applied to plant identification. Information Sciences, 178(12):2722 – 2733, 2008.
[8]. J.-X. Du, X.-F. Wang, and G.-J. Zhang. Leaf shape based plant species recognition. Applied Mathematics and Computation, 185(2):883 – 893, 2007.
[9]. Ming-Kuei Hu, “Visual Pattern Recognition By Moment Invariants” Ire Transactions On Information Theory.
[10]. Krishna Singh, Indra Gupta, Sangeeta Gupta, SVM-BDT PNN and Fourier Moment Technique for classification of Leaf, International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 3, No. 4, December, 2010.
[11]. International Journal of Modern Engineering Research (IJMER) www.ijmer.com Vol.1, Issue.2
[12]. Sandeep Kumar.E, “Leaf Color, Area And Edge Features Based Approach For Identification Of Indian Medicinal Plants” Indian Journal Of Computer Science And Engineering .
[13]. Abdolvahab Ehsanirad And Sharath Kumar Y. H. ,“Leaf recognition for plant classification using GLCM and PCA methods” Oriental Journal of Computer Science & Technology, Vol,3(1),31-36, 2010.
[14]. Hong Fu and Zheni Chi, “A two-stage approach for leaf vein extraction”, Proceedings of International Conference on Neural Networks and Signal Processing, Vol. 1, Pp. 208–211, 2003.
[15]. Zalikha Zulkifli, Puteh Saad and Itaza Afiani Mohtar, “Plant leaf identification using moment invariants & General Regression Neural Network”, 11th International Conference on Hybrid Intelligent Systems (HIS), Pp. 430–435, 2011.

[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]. Zheru Chi, Li Houqiang and Wang Chao, “Plant species recognition based on bark patterns using novel Gabor filter banks”, Proceedings of International Conference on Neural Networks and Signal Processing, Vol. 2, Pp. 1035 – 1038, 2003.
[18]. R. Yusof, N.R. Rosli and M. Khalid, “Using Gabor Filters as Image Multiplier for Tropical Wood Species
[19]. Recognition System”, 12th International Conference on Computer Modelling and Simulation (UKSim), Pp. 289–294, 2010.
[20]. Marzuki Khalid, Rubiyah Yusof,, Mohd. Khairuddin and Anis Salwa, “Tropical wood species recognition system based on multi-feature extractors and classifiers”, 2nd International Conference on Instrumentation Control and Automation (ICA), Pp. 6–11, 2011.
[21]. K.J. Runtz, “Electronic recognition of plant species for machine vision sprayer control systems”, IEEE Western Canada Conference on Computer, Power and Communications Systems in a Rural Environment, Pp. 84-88, 1991.
[22]. Lei Zhang, Jun Kong, Xiaoyun Zeng and Jiayue Ren, “Plant Species Identification Based on Neural Network”, Fourth International Conference on Natural Computation (ICNC '08), Vol. 5, Pp. 90–94, 2008.
[23]. A.H. Kulkarni, Dr. H.M.Rai, Dr.K. A. Jahagirdar and P.S.Upparaman, “A Leaf Recognition System for Classifying Plants Using RBPNN and pseudo Zernike”, International Journal of Latest Trends in Engineering and Technology (IJLTET).
[24]. Javed Hossain, M. Ashraful Amin, “Leaf Shape Identification Based Plant Biometrics”, Proceedings of 13th International Conference on Computer and Information Technology 23-25 December, 2010, Dhaka, Bangladesh.
[25]. Jyotismita Chaki, Ranjan Parekh, “Designing An Automated System For Plant Leaf Recognition”, International Journal Of Advances In Engineering & Technology, Jan 2012, ISSN: 2231-1963.
[26]. Anant Bhardwaj, Manpreet Kaur and Anupam Kumar, “Recognition of plants by Leaf Image using Moment Invariant and Texture Analysis”, International Journal of Innovation and Applied Studies ,ISSN 2028-9324 Vol. 3 No. 1 May 2013, pp. 237-248
[27]. M K HU, “Visual pattern recognition by moment invariants,” IRE Trans. Info Theory, vol IT-8, pp. 179-187, Feb. 1962.

