IFERP
connecting engineers... developing research

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 4, Issue 6, June 2017

Plant identification system using leaf feature Plant

[1] Ms. Vinutha B T, [2] Arvind Kumar, [3] Chauhan Akash, [4] Nithin M E, [5] Prashanth G D
[1] Asst. Prof. Dept. of CSE, RR Institute of Technology, Bangalore, Karnataka
[2][3][4][5] UG Student, Dept. of CSE, RR Institute of Technology, Bangalore, Karnataka

***Abstract*— From the beginning of earth survival the plant life begin. Plant affecting in flora and fauna. 70% of plant life found in ocean. Year passes revolutionary happen and human life came to survival. Due to that evolution happen in human life. Man started study on plant life. They got the outcome of their study that plants are being used as survival of human being. After that they started study more about plant and they get to know more about plant and there uses such as it can be used as medicine & etc. Human are totally depends on plant life on the daily bases. But the major occur that extinction of plant. Due to this, it effects on earth by increasing rate of pollution and Ozone layer is getting decreases and UV ray is radiating.**

Many research happened on plant and it terms as Botany. Here also, we are going to propose the system for identification of a leaf. We are going to extract the features of leaf such as perimeter, area, eccentricity, centroid & etc.

It is difficult to recognize whether it is medicinal plant or not. But using a MATLAB it is easy to identify the leaf and their features. In this MATLAB we can capture a real time leaf image using a high pixel camera. After capturing a image it will go under different process. According to DATABASE the leaf id generates and provides its basics features.

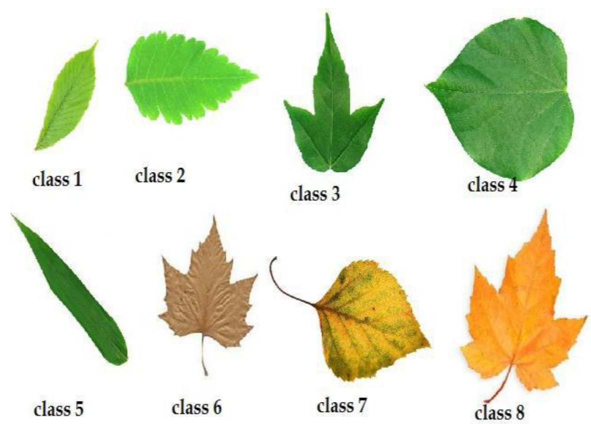
***Index Terms*—Leaf classification, color features, image processing, shape and texture features.**

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.

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 [2][3]. 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. [4] 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 [5] 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 [6], 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.

II. CLASSIFICATION OF PLANTS LEAF'S



ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 4, Issue 6, June 2017

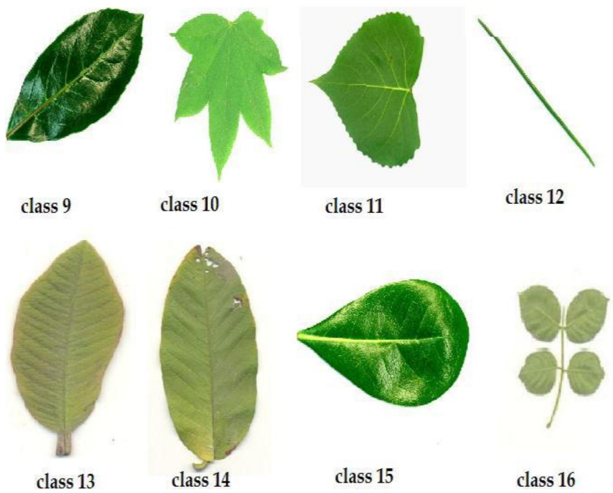


Fig. 1: Classification of Plant leaf's

III. 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 'region props' in mat lab 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. Equivalent Diameter: It is the diameter of a circle with the same area as the region. It is Computed as sqrt (4*Area/pi).

IV. IDENTIFICATION METHODS

There are several methods which can be used to identify plant leaves and other materials. Some of the widely used methods for identification are: Spectroscopy, Chemical identification and Optical identification. Spectroscopy measures the interaction of the molecules with electromagnetic radiation which can be used for the identification of leaves. Chemical identification involves identification of key chemicals in a particular plant. Each plant and especially medicinal plants, which are the subject of this study, have particular chemicals which impart the useful characteristics to the plant and these can be identified using analytical chemistry. Our proposed algorithm uses optical method. We will be using Digital image processing for the identification of leaves. In this method, we will be using a computer algorithm to process the leaf images. Digital image processing is superior to analog image processing since we will be able to avoid signal distortion and noise during the processing. The steps involved in optical method are feature extraction, pattern recognition and classification in the same order. We are using a computer program for classification to increase the reliability of classification and decrease the time required for classification. A computer program also eliminates the scope for human error.

V. TEXTURE FEATURES

An image texture is a set of metrics calculated in image processing designed to quantify the perceived texture of an image. Image texture gives us information about the spatial arrangement of color or intensities in an image or selected region of an image. Image textures can be artificially created or found in natural scenes captured in an image. Image textures are one way that can be used to help in segmentation or classification of images. For more accurate segmentation the most useful features are spatial frequency and an average grey level. To analyze an image texture in computer graphics, there are two ways to approach the issue: Structured Approach and Statistical Approach. 1. Edge Detection: The use of to determine the number of edge pixels in a specified region helps determine a characteristic of texture complexity. After edges have been found the direction of the edges can also be applied as a characteristic of texture and can be useful in determining patterns in the texture. These directions can be represented as an average or in a histogram. Consider a region with N pixels. the gradient-based edge detector is applied to this region by producing two outputs for each pixel p: the gradient magnitude Mag(p) and the gradient direction Dir(p). The edginess per unit area can be defined by

$$F_{edgeness} = \frac{|\{p|Mag(p) > T\}|}{N} \quad \text{…………… (1)}$$

for some threshold T.

To include orientation with edginess we can use histograms for both gradient magnitude and gradient direction. Let Hmag(R) denotes the normalized histogram of gradient magnitudes of region R, and let Hdir denote the normalized histogram of gradient orientations of region R. Both are normalized according to the size NR Then

$$F_{magdir} = (H_{mag}(R), H_{dir}(R)) \quad \text{…………….(2)}$$

is quantitative texture description of region R. 2. Co-occurrence Matrices: The co-occurrence matrix captures numerical features of a texture using spatial relations of similar gray tones. Numerical features computed from the co-occurrence matrix can be used to represent, compare, and classify textures. The following are a subset of standard features derivable from a normalized co-occurrence matrix:

$$Angular\ 2nd\ Moment = \sum_i \sum_j p[i,j]^2$$

$$Contrast = \sum_{n=0}^{Ng-1} n^2 \left\{ \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p[i,j] \right\}, \text{ where } |i-j| = n$$

$$Correlation = \frac{\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} (ij)p[i,j] - \mu_x \mu_y}{\sigma_x \sigma_y}$$

$$Entropy = -\sum_i \sum_j p[i,j] ln(p[i,j]) \quad \text{………………(3)}$$

where is the entry in a gray-tone spatial dependence matrix, and Ng is the number of distinct gray-levels in the quantized image. One negative aspect of the co-occurrence matrix is that the extracted features do not necessarily correspond to visual perception.

3. Laws Texture Energy Measures:

Another approach to generate texture features is to use local masks to detect various types of textures. Convolution masks of 5x5 are used to compute the energy of texture which is then represented by a nine element vector for each pixel. The masks are generated from the following vectors:

L5 = [+1 +4 6 +4 +1] (Level)

E5 = [-1 -2 0 +2 +1] (Edge)

S5 = [-1 0 2 0 -1] (Spot)

W5 = [-1 +2 0 -2 +1] (Wave)

R5 = [+1 -4 6 -4 +1] (Ripple)

Sixteen 5x5 convolution masks are then generated as the outer product of pairs of the above masks as follows:

L5L5

E5E5

S5S5

R5R5

L5E5

E5L5

L5S5

S5L5

L5R5

R5L5

E5S5

S5E5

E5R5

R5E5

S5R5

R5S5

4. Autocorrelation and Power Spectrum:
The autocorrelation function of an image can be used to detect repetitive patterns of textures.

VI. TEXTURE SEGMENTATION

The use of image texture can be used as a description for regions into segments. There are two main types of segmentation based on image texture, region based and boundary based. Though image texture is not a perfect measure for segmentation it is used along with other measures, such as color, that helps solve segmenting in image.

1. Region Based:
Attempts to group or cluster pixels based on texture properties.

2. Boundary Based:
Attempts to group or cluster pixels based on edges between pixels that come from different texture properties.

VII. 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[7]. 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.[8]

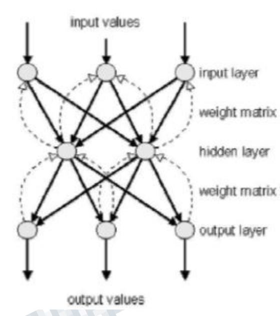


Fig. 3: Input-Output values.

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.

3. Moment Invariant Method (MI):
Moments and functions of moments have been extensively employed as invariant global features of images in pattern recognition[9]. 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 [8].

4. 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. Speech in the early 1990s.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

VIII. CONCLUSIONS

The work presented in this paper is an attempt to perform automatic recognition of medicinal plants, as well as to analyse the statistical nature of the image features used for recognition. Ten plant species including plants belonging to the same families were used for the experimental analysis. Compared to our previous work, few more medicinal plant species were considered and it was proved that the leaf features are Gaussian distributed, which served as a useful result in accurate classifier design by calculating the precise decision boundaries between the classes. This work is a contribution towards serving horticulture in India and an attempt towards a faster and better means of identifying medical plant species.

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]. 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.

[3]. 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 Computer Vision (ECCV), pages 116–129, 2008.

[4]. 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.

[5]. 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.

[6]. International Journal of Modern Engineering Research (IJMER) www.ijmer.com Vol.1, Issue.2

[7]. Ming-Kuei Hu, "Visual Pattern Recognition By Moment Invariants" Ire Transactions On Information Theory.Ming-Kuei Hu, "Visual Pattern Recognition By Moment Invariants" Ire Transactions On Information Theory.

[8]. 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.

[9]. Javed Hossain, M. Ashrafull Amin, "Leaf Shape Identification Based Plant Biometrics", Proceedings of 13th International Conference on Computer and Information Technology 23-25 December, 2010, Dhaka, Bangladesh.

[10]. Jyotismita Chaki, Ranjan Parekh, "Designing An Automated System For Plant Leaf Recognition", International Journal Of Advances In Engineering & Technology, Jan 2012, ISSN: 2231-1963.