Leaf Recognition Based on Color, Shape and GLCM Features

Naveena M[1], G Hemantha Kumar[2], Vidyashanakara[3], Pavithra B S[4]

[1, 2, 3] *DoS in Computer Science, University of Mysore, Mysuru, India*
[4]*IIHR, Bangalore, India*

***Abstract*:-Getting to know the details of plants growing around us is of great importance medicinally and economically. Conventionally, plants are categorized mainly by taxonomists through investigation of various parts of the plant. However, most of the plants can be classified based on the leaf shape and associated features. Image processing techniques are used to extract leaf color, shape and GLCM features such as mean, standard deviation, homogeneity, correlation, aspect ratio, width ratio, apex angle, apex ratio, base angle, centroid deviation ratio, moment ratio and circularity. Under the current research, 450 leaves of 9 kinds of plants were collected. Out of these, 270 leaves were trained. The 180 testing samples were recognized with 92% accuracy; even without considering types of leaf margins, vein and removal of the petiole.**

***Keywords*—aspect ratio; apex ratio; base angle; energy; homogeneity.**

I. INTRODUCTION

For all forms of life, plants form the basic food staples, and this is just one reason why plants are important. They are the major source of oxygen and food on earth since no animal is able to supply the components necessary without plants. The fish we eat consume algae and the cattle we eat as beef feed on grass, so even if you are not a fan of salads, your food source relies on plants. Plants also provide animals with shelter, produce clothing material, medicines, paper products, reduce noise levels and wind speed, reduce water runoff and soil erosion. Coal is also produced from plant materials that were once alive. Plants are very much significant component of ecosystem. Plants can be used as foodstuff, in medicines and in many industries for manufacturing various products. Identifying those helps ensure the protection and survival of all natural life. Plant identification can be performed using many different techniques. As plant leaves are more readily available, it is efficient to identify and classify plants by using their leaves. Plant classification by using leaves requires different biometric features. All that gives plants its important role in life on earth. For example, as natural resource managers, they must understand what they manage, and plant recognition is a key component of that understanding. The ability to know, or recognize plants allows them to assess many important rangeland or pasture variables that are critical to proper management: range condition, proper stocking rates, forage production, wildlife habitat quality, and rangeland trend, either upward or downward. Natural resource managers, especially those interested in grazing and wildlife management must be able to evaluate the presence or absence of many plant species in order to assess these variables. Plants exist everywhere we live, as well as places without us. Plants are mostly identified by taxonomists and the process is usually lengthy. Plant features that help in identifying a plant are fruit, seed, leaf, flower, root, stem, etc. India is an agriculture country where in about 70% of the population depends on agriculture. Farmers have wide range of diversity to select suitable Fruit and Vegetable crops. Leaves are used in herbal medicine, soups, stews, meat, seafood and vegetable dishes, etc... So, Getting to know the details of plants growing around us is of great importance medicinally and economically.

Leaves play an important role in identification of a plant due to its availability nearly throughout the year.

II. ABOUT THE PROBLEM

Till date a large volume of work has been done in the area of leaf recognition. Many approaches which are discussed in literature survey of the report are used to recognize the leaves of different plants. Among all these recognition work on leaves of plants is not so common to that of Optical character recognition and Face recognition. From the literature it came out that color shapes and GLCM features are not used to leaf recognition. This motivates me to put effort to develop a leaf recognition system based on color shapes and GLCM features.

II.a) Approaches

The RGB color image of leaf has been obtained from different plants. Before the main leaf recognition, a set of preprocessing operations has been performed on the RGB leaf images. Preprocessing includes following operations like normalization, gray scale conversion, binary conversion and we apply median filter to remove noise. Further to extract the shape features such as Aspect ratio, Width ratio, Apex angle, Apex ratio and Base angle. To extract GLCM properties such as Contrast, Correlation, Energy and Homogeneity we need Grey scale image using the gray scale image we can extract the GLCM properties. Then with the help of designed classifier, leaves are recognized.

III. MOTIVATION

Plant classification based on leaf identification is becoming a popular trend. Each leaf carries substantial information that can be used to identify and classify the origin or the type of plant. In medical perspective, images have been used by doctors to diagnose diseases. Some experiments have yielded many success stories in the lab, but some approaches have failed miserably when tested in the real world. This happens because researchers may have ignored the facts that the real world sampling may not have the luxury and complacency as what they may have in the lab.

Finally, an extra motivation, from an application point of view, is to develop an

application of the proposed techniques through the new technologies and devices, such as mobiles and tablets.

IV. CHALLENGES

Identification of leaf accurately only by the consideration of shape, color features of the leaf image without considering the leaf veins itself is a main challenge.

Image condition is also important for recognizing the image, it deals with intensity, resolution, camera lighting, background, characteristics of image capturing device and distance between camera and person, plays an important role in the process of face detection.

II.b) Applications

Proposed approach in plant classification and recognition that not only applicable in the real world, but also acceptable in the lab. Plants have many uses in industry, medicine, and foodstuff production. A computer-based plant classification system can use various characteristics of plants such as leaves, flowers, fruits, branching styles, and outlooks. An easier and accurate way is using leaves to identify plants.

V. PROPOSED METHODOLOGY

We used three groups of features from leaf images in this project. The first group of features is based on color values of leaf images, second group of features is based on shapes of leaf images, and the third group is based on gray level co-occurance matrix (GLCM). KNN classifier is used for classification.

Main features	**Sub-features**
Shape	Aspect ratio, circularity, apex angle, base angle, apex ratio, width ratio
Color	Mean , Variance , Range of Red, Green , Blue (RGB) and Hue, Saturation, Value(HSV)
GLCM	
	Contrast, Co relation, Energy, Homogeneity

Table : Represents Main feature and Sub-features

Color Features:

Separate RED, GREEN, BLUE, HUE, SATURATION, VALUE components of the normalized Image. Then find out the Mean, Range, Standard deviation and sum of the range for RED, GREEN, BLUE, HUE, SATURATION, VALUE component of the normalized image.

Mean Formula

$$\mu = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N} p_{xy}$$

Range Formula

Range = max2-min2

Standard deviation formula

$$\sigma = \sqrt{\frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}(p_{xy}-\mu)^2}$$

Shape Features

We used 5 fundamental features of leaf

1. Major axis length (L):- The distance between base and apex.

2. Max width (W):- Maximum width of the leaf.

3. Width at 0.5L:-Width of the leaf at 50% of the major axis.

4. Apex angle:-Angle at apex between leaf edges on width line at 75% of the major axis(measured from base).

5. Apex angle at 0.9L:-Angle at apex between leaf edges on width line at 90% of the major axis.

6. Base angle:- Angle at base between leaf edges on width line at 25% of the major axis.

Final Shape Features:

•**Aspect ratio:** The ratio of maximum width to major axis length.

•**Width ratio:** The ratio of width at half of major axis to maximum width.

•**Apex angle:** The angle at apex between leaf edges on width line at 3/4th of the major axis.

•**Apex ratio:** The ratio of the angle made at the apex by width at 9/10th of major axis to angle at 3/4 th of major axis.

•**Base angle:** The angle at base between leaf edges on width line at 1/4th of the major axis.

•find the distance of every plotted pixel to the center of the image

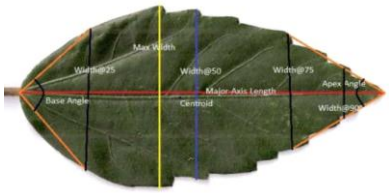


Fig: shape features

GLCM Properties

•Contrast

Returns a measure of the intensity contrast between a pixel and its neighbour over the whole image. Contrast is 0 for a constant image.

•Correlation

Returns a measure of how correlated a pixel is to its neighbor over the whole image. Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is NaN for a constant image.

•Energy

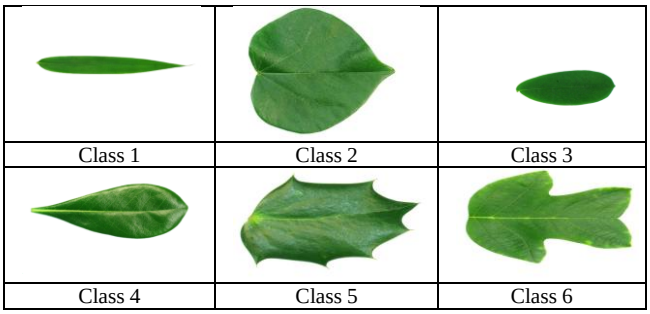
Returns the sum of squared elements in the GLCM. Energy is 1 for a constant image.

•Homogeneity

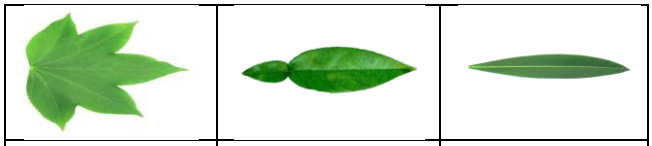
Returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. Homogeneity is 1 for a diagonal GLCM.

VI. DATASET

We have collected totally 450 images. The dataset contains 9 classes, each class contains 50 images. We have taken only the healthy leaf images of pre-matured and matured. Class one contain 50 (30 training + 20 testing) leaves, Class two contains 50 leaves, so on..,Samples images of the collected dataset are shown in



Class 1	Class 2	Class 3
Class 4	Class 5	Class 6



Class 7	Class 8	Class 9

We performed three different experiments with various sets of features. In the first experiment, we made classification with the shape features. In the second experiment, we used the shape and the color features together in classification. In the last experiment, we used all shape, color features (both RGB and HSV set of color features), GLCM together.

KNN Classifier

Neighbors-based classification is a type of instance-based learning or non generalizing learning: it does not attempt to construct a general internal model, but simply stores instances of the training data. Classification is computed from a simple majority vote of the nearest neighbors of each point: a query point is assigned the data class which has the most representatives within the nearest neighbors of the point.

Class = knn classify (Sample, Training, Group, k) enables you to specify k, the number of nearest neighbors used in the classification. Default is 1.

Sample: Matrix whose rows will be classified into groups. Sample must have the same number of columns as Training.

Training: Matrix used to group the rows in the matrix Sample. Training must have the same number of columns as Sample. Each row of Training belongs to the group whose value is the corresponding entry of Group.

Group: Vector whose distinct values define the grouping of the rows in Training.

K: The number of nearest neighbors used in the classification. Default is 1.

Table : Indicates the accuracies with features

Class	Accuracy with only Color features	Accuracy with Color + shape features	Accuracy with Color + shape + GLCM features
Class1	73	81	89
Class2	84	85	86
Class3	78	82	91
Class4	86	90	88
Class5	71	76	84
Class6	89	92	91
Class7	74	75	79
Class8	86	90	92
Class9	82	88	92

VII. Conclusion

Image processing techniques are used for extracting the morphological parameters that are having some significance and effect on the classification of the leaves. This project gives the types of leaf features that should be extracted, external factors the must be considered before the extraction process, types of extraction and classification method that can be used for plant recognition and classification. In otherwords, the results of this project can be used as a specification of leaf features that must be considered for plant recognition and classification.

The blue band of the leaf may be used as grayscale image rather than composite image created using the standard color-gray conversion formula. The new parameter apex ratio shows how much conical the apex of a leaf is. The same species with clearly different shape feature may be processed as separate sample type. The leaflet of a compound leaf can also be considered as a sample for recognizing a plant.

VIII. Future work

Next step in the future work, different classifiers will be tested based on our dataset and the results will be recorded. Only the better classifier will be used in our research. However, we may have to consider images that contain many leaves in order to test the ability of the classifiers and we are looking to consider leaf margins, veins for better identification of confusing leaves.

•Removal of shadow and complex background

•Automatic identification of base and apex

REFERENCES

[1]. Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu xuan Wang, Yi fan Chang, and Qiaoliang Xiang. A leaf recognition algorithm for plant classification using probabilistic neural network. IEEE 7th international Symposium on Signal Processing and Information Technology, 2007.

[2]. Uluturk C and Ugur A. Recognition of leaves based on morphological features derived from two half-regions. In Innovations in Intelligent Systems and Applications (INISTA), 2012 International Symposium on, pages 1- 4. IEEE, 2012.

[3]. Xiao Gu, Ji-Xiang Du, and Xiao-Feng Wang. Leaf recognition based on the combination of wavelet transform and gaussian interpolation. LNCS 3644, pages 253-262. Proceedings of International Conference on Intelligent Computing 2005, 2005.

[4]. Mishra, S.K. Maurya, R.K. Singh, and AK Misra. A semi automatic plant identification based on digital leaf and flower images. In Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on, pages 68-73. IEEE, 2012

[5]. Odemir Martinez Bruno, Rodrigo de Oliveira Plotze, Mauricio Falvo, and Mriode Castro. Fractal dimension applied to plant identification. Information Sciences,178(12):2722- 2733, 2008.

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

[7]. Smita Naikwadi ―ADVANCES IN IMAGE PROCESSINGFOR DETECTION OF PLANT DISEASES‖ International Journal of Application or Innovation in Engineering & Management(IJAIEM) WebSite:www.ijaiem.org Volume 2, Issue 11, November 2013.

[8]. Amin A.M and A. I. Khan A.I, "One-shot Classification of 2-DLeaf Shapes Using Distributed Hierarchical Graph Neuron (DHGN)Scheme with k-NN Classifier," Procedia Computer Science, 24, 2013,pp.84-96

[9]. Samuel E. Buttrey et.al. ―Using k-nearest-neighbor classification in the leaves of a tree‖ Computational Statistics &DataAnalysis40(2002)27 –37www.elsevier.com/locate/csda.

[10]. Amlekar Manisha, Manza R.R, Yannawar Pravin, GaikwadB.P,(2013), Image data mining for classifying leaf dimension biometric features of leaf shape using KNN classification technique, CMS.

[11]. Aurangabad. Beghin T., Cope J. S., Remagnino P. and Barman S.,(2010), Shape and texture based plant leaf Classification, ACIVS, 2,345–353.

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