Computer Vision Based Feature Extraction of Leaves for Identification of Medicinal Values of Plants

D Venkataraman
Department of Computer Science and Engineering,
Amrita School of Engineering, Coimbatore,
Amrita Vishwa Vidyapeetham,
Amrita University,
India.
d_venkat@cb.amrita.edu

Mangayarkarasi N
Department of Computer Science and Engineering,
Amrita School of Engineering, Coimbatore,
Amrita Vishwa Vidyapeetham,
Amrita University,
India.
cb.en.p2cvi15004@cb.students.amrita.edu

Abstract— **Plants are considered as one of the greatest assets in the field of Indian Science of Medicine called Ayurveda. Some plants have its medicinal values apart from serving as the source of food. The innovation in the allopathic medicines has degraded the significance of these therapeutic plants .People failed to have their medications at their door step instead went behind the fastest cure unaware of its side effects. One among the reasons is the lack of knowledge about identifying medicinal plants among the normal ones. So, a Vision based approach is being employed to create an automated system which identifies the plants and provides its medicinal values thus helping even a common man to be aware of the medicinal plants around them. This paper discusses about the formation of the feature set which is the important step in recognizing any plant species.**

Keywords— SVM, Image pre processing, Probabilistic Neural Network, Principal Component Analysis, Feature Extraction

I. INTRODUCTION

The plants which are around us play a major part in framing our ecosystem. Some gives us edible eatables either above or beneath the ground whereas there are a wide variety of plants that has found its use in Indian system of medicine called Ayurveda. These plants are different from the normal ones. In earlier days, people were good enough to identify the medicinal aspects of these plants in curing various diseases. These plants were the ones that normally grow in our backyards or the ones that we find along the roadsides. As the days pass it is becoming difficult for the people to identify the existence of the medicinal plants. Many are unaware of these plants. So, to identify a plant first we consider the leaves of that plant to classify them. Leaves can be classified based on various features like texture, shape and color .The texture based feature classification has been discussed in a paper [1] where a statistical approach has been used which gives the quantitative measure of the pixel arrangements in a region. So, the GLCM method has been used and the dissimilarity between the leaves has been found and leaves are classified. This algorithm works on different herbal leaves namely Thulasi, Omavalli, Neem, Vana Thulasi, Thudhuvalai and nochi. In some cases color and shape features [2] has been used to get the feature vectors from where the classification is done. There is various classification algorithms used to classify the plants based on the feature vectors. Some of the papers discuss about Support Vector Machine (SVM) [3] and Artificial Neural Network [4] algorithm. E.Sandeep Kumar and Viswanath Talasila [5] has discussed about the Gaussian distribution of the leaf features that serves as the effective classifier. In this paper, survey on various algorithms like Probabilistic Neural Network, Support Vector Machine, Principal Component Analysis and the texture base analysis used to detect a plant is discussed further.

II. EXISTING METHODOLOGY

A. Pre Processing

The leaves are considered as a dominant feature for identifying a plant type. The digital image of the leaf part of any plant is given as an input data. This image undergoes preprocessing steps in order to remove any kind of external noises present in an image. The main idea of preprocessing is to enhance the image details so that features are clearly found for further processing. The noise removed leaves re then converted from color to grayscale image which will be easy for feature extraction process. The contour of the leaf is then detected using the edge detectors.

B. Feature Extraction

Each leaf possess unique feature that makes it different from the other. In general features are classified based on shape [3], color and textures [1]. Features are described in brief below. The *Fig.1* shows the classification of features based on the categories mentioned.

978-1-5090-0612-0/16/$31.00 ©2016 IEEE

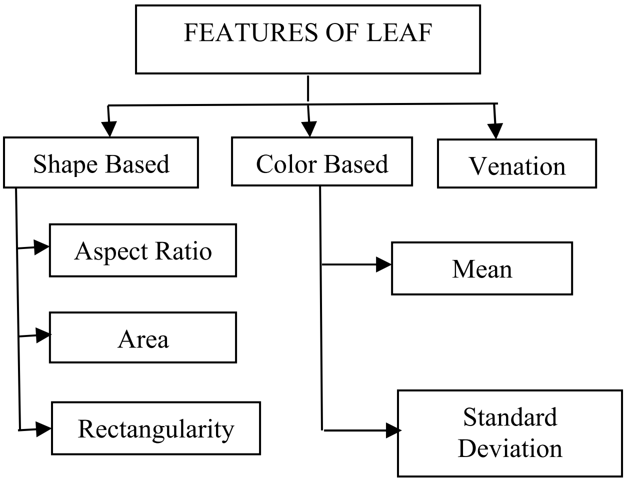

Fig. 1. Types of features

1) Shape Features:

There are different shape features defined based on the geometry of the leaves. The length of the leaf is found by taking the Euclidean distance [3] between the two tip points on the either side long the axis whereas breadth corresponds to the length of minor axis as in *Fig. 2.*

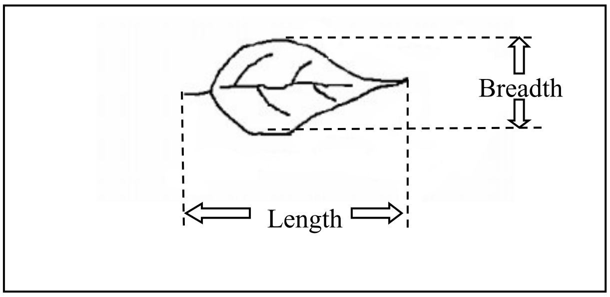

Fig.2. Section of leaf

From the length and breadth of the leaf the aspect ratio is found which is the ratio between the length and breadth of the leaf. It is given as in (1)

$$\text{Aspect Ratio} = \frac{\text{Length of the leaf}}{\text{Breadth of the leaf}} \tag{1}$$

- The area is calculated by initially fining the area of one pixel as in (2).

$$\text{Area} = \text{Area of pixel} * \text{Total no : of pixels present in a leaf} \tag{2}$$

- Perimeter of the leaf is given by the count of the pixels having the leaf margin.
- Rectangularity describes the similarity between a leaf and a rectangle. It is defined as :

$$\text{Rectangularity} = \frac{L * W}{A} \tag{3}$$

Where L is the length, W is the width and A is the area of the leaf .These features correspond to the shape of the leaf.

2) Color Features:

The color features include mean (4) and standard deviation (5) where mean is the sum average of the total pixel in the leaf given as:

$$Mean\ \mu = \frac{\sum\sum I(x, y)}{M * N} \tag{4}$$

$$\text{Standard deviation } \sigma = \sqrt{\left(\frac{1}{M * N}\sum\sum(I(x, y) - \mu)^2\right)} \tag{5}$$

3) Venation:

The leaves can be better identified by its veins. These veins are unique for each plant species. [5]These can be extracted by applying the morphological operations like opening by using the structuring element. So, that the background information is subtracted and only the vein patterns appear.

C. Probablistic Neural Network

Probabilistic neural networks is a type of classification algorithm that is derive from the from Radial Basis Function (RBF) Network. The major advantage of PNN is its time complexity. It takes lesser time to train the data.[5]The working of PNN includes three layers namely input layer, Radial Basis Layer and competitive layer. The input layer is the first layer through which the input vectors are fed. As the input is given the distance of the input vectors are calculate with respect to the vectors in the training dataset. A vector value [5] is got as a result that shows how similar s the input vector to that of training data. The values of each input class are summed in the second layer thus giving an output vector. The distance between the row weight vectors of the weighted matrix and the input vectors re calculated by the Radial Basis Layer which then scales these distance values nonlinearly. Here the weight values [5] are assigned instead of training. The new vectors can be added into the weight matrix without altering the existing ones. The competitive layer is the last layer that decides to which class the input data belongs to base on the probabilities. The one whose probability is maximum takes the value 1 and the remaining classes takes 0.

D. Support Vector Machine

Support Vector Machine is a classifier that separates the input data non-linearly by mapping it to a high dimensional space. SVM is a very efficient classifier and is mainly used for solving binary classification problems. There are two main steps involved in SVM classification technique namely,

- Transforming input data to higher dimensional space by using Kernel function
- Construction of OSH(Optimal Separating Hyper Plane)

The feature vector obtained from the boundary or shape of the leaf is used to classify a plant leaf. To find the OSH, SVM increases the margin among the classes. Support vectors are the data vectors in the transformed space that are near the constructed OSH plane. Generally, SVM derives a function to discriminate the input data into two classes [2]. With the help of a non linear transformation which is regularization parameter dependant, we move the input vectors into a high-dimensional feature space, where a linear separation is employed. This is then configured as a nonlinear support vector classifier by replacing the inner product (x, y) with a kernel function K (x, y), as shown below

$$F(X) = \operatorname{sgn} \sum_{i=1}^{l} a_j y_j K(X_i X) + b \qquad (6)$$

where f(x) determines the membership of x. We consider that normal subjects are labeled as -1 and the other subjects as +1. SVM is said to have two layers. The initial layer selects the basis K (x_i x), i=1, 2… N from the defined kernels, followed by the second layer with the help of which a linear function is constructed in space [6]. Various types of learning machines can be constructed by using different kernel functions. The advantage is that, SVM holds the property of simple geometric interpretation and thus providing a sparse solution. Also the independency of computational complexity with dimensionality of input space holds as strength to this technique. The only underlying limitation is that, SVM uses many support vectors from the training data to classify the input test data.

E. *Principal component analysis*

Principal component analysis is a technique used to reduce the variance and is also called as variable reduction technique. In this method [5], the processed data is projected from a higher dimensional space to a lower dimension. PCA generates a set of principal components each of which component is a linear combination of optimally – weighted variables. Since these components are orthogonal to each other, redundancy is reduced to a greater extent. Mathematically, the data is transformed to a new coordinate system [5] where the first coordinate will have the data with first highest variance and the second coordinate holds the data with second highest variance and so on. At this stage, each coordinate is called as a Principal component. The variability of the data will be reduced to a great extent even with the help of small number of principal components. PCA reduces the dimensions [2] at a large scale by still preserving the randomness or the spread of the data in the high dimensional space. But the interpretation of principal components is not easy because PCA does not consider the class label of the feature vector and its class seperability. This stands as a drawback of PCA technique.

III. PROPOSED DESIGN

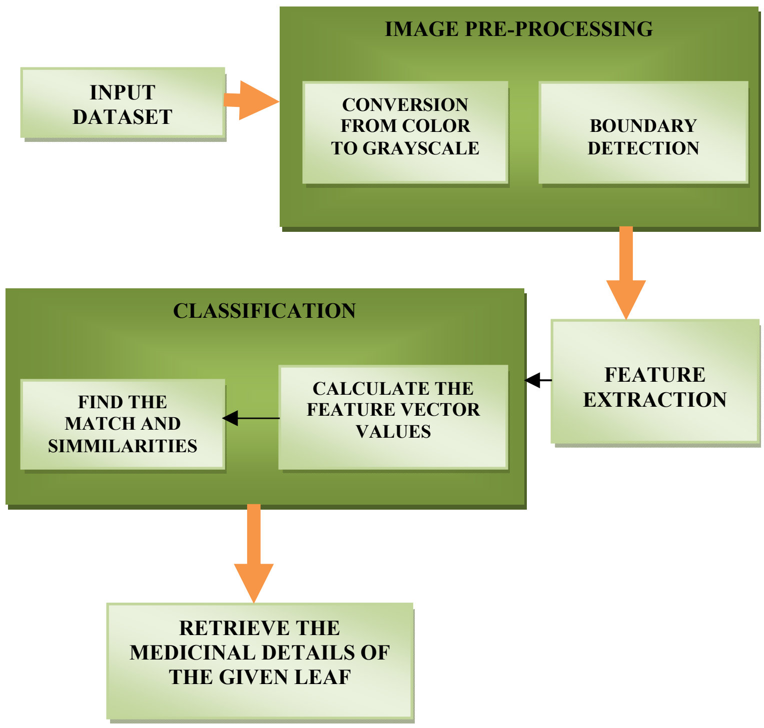

Fig.3.Architectural design

Initially the leaf image is given as an input to the system which is then converted to gray scale image. This gray image is used for edge detection which depicts the boundary of the leaf .Having this as the input varies geometric shape based features are extracted that forms the feature set. This feature set is then compared with the values in the database i.e. the trained results and then matched values are classified as a particular species. Once classified the medicinal values of that plant species is retrieved and displayed to the user.

A. *Defining a Feature Set*

The detection of the boundary for the leaf specimen involves the binarization of the leaves. The erosion operation is applied in order to shrink the boundaries of the foreground pixel region. The image subtraction is done between the binary image and the eroded one which neglects the other intensity values to black and highlights the boundary alone with white pixel values. The output obtained is as show in *Fig. 4*. Here we consider a neem leaf for our process initially.

.

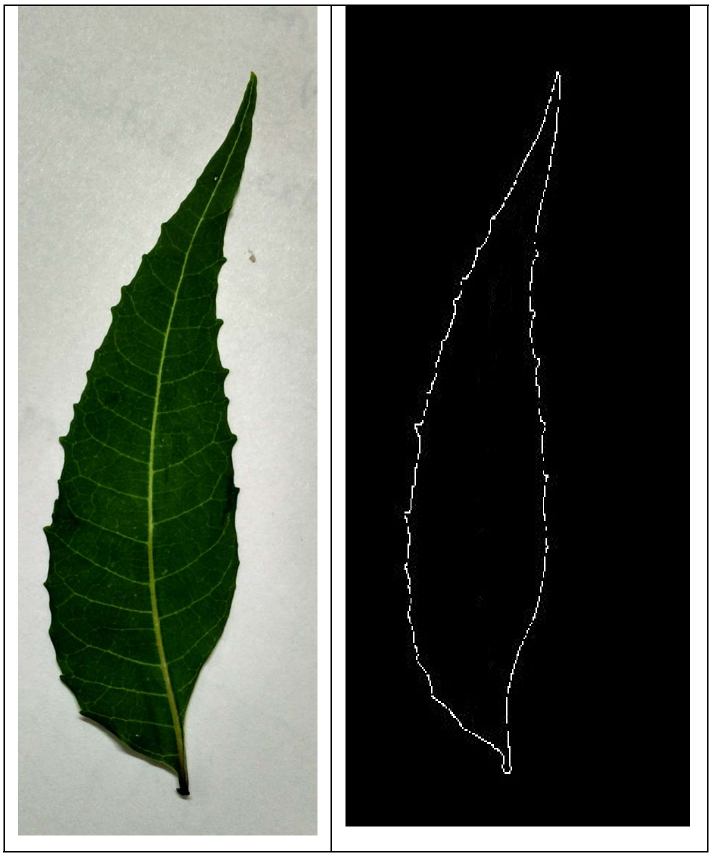

Fig.4. Input Neem leaf and Boundary detected image

As we know the shape of the leaves varies from one species to another, but the size varies from one species to another and also within its own species itself as the plant grows. So, this is taken as the essential point which involves the measure of the length and the breadth of each leaf as described in the above section. This serves as the data for calculating the Aspect Ratio, Centroid, Area, Perimeter and Roundness of the leaf.[11] Each feature has its own property that defines a leaf structure of a group. We have considered five different sizes of the neem leaf for which the feature set is defined. Not all leaves are completely elliptical or circular so this is described using the Roundness value. This shows how round a pixel is as discussed in previous section.This varies with different plants since some leaves are narrow and others are broad. This would be one among the unique features that recognizes a leaf.

IV. RESULTS AND DISCUSSION

The features are extracted for five different leaf samples of neem and the feature set is tabulated as in Table I. Based on the results the findings made include:

- Aspect ratios of a particular leaf do not vary as the size of the leaf varies. It is distinct for one particular species.
- Roundness value of the leaf varies with the size of the leaves. So a range should be fixed for each species of plant.

TABLE I.

FEATURE SET OF DIFFERENT NEEM LEAF SAMPLES

Aspect Ratio	**Area**	**Perimeter**	**Roundness**
1.7770	9964395	1.3257e+04	0.7125
1.7770	9964543	1.3149e+04	0.7243
1.7770	9964540	1.3151e+04	0.7240
1.7770	9964544	13148	0.7243
1.7770	9964544	13148	0.7243

V. CONCLUSION AND FUTURE WORK

The features can be calculated for different types of herbal leaves and can be stored in the database which will be the trained result values. The future work will be on framing a robust algorithm and make it trained with leaf datasets so that when the user gives a dataset in the real-time ,the algorithm matches it with the trained dataset. It then gives to which plant it belongs to and the medicinal values are displayed to the user.

ACKNOWLEDGMENT

The authors would like to extend the heartfelt gratitude to the faculty-in-charge of Amrita-Cognizant Innovation Lab, Department of Computer science and Engineering, Amrita school of Engineering, Coimbatore for the support extended in carrying out this work.

REFERENCES

[1] T. Sathwik, R. Yasaswini, Roshini Venkatesh and A. Gopal,"Classification of Selected Medicinal Plant Leaves Using Texture Analysis", 4th ICCCNT, July 4 -6, 2013.

[2] Shitala Prasad, Krishna Mohan Kudiri, and R.C. Tripathi ," Relative Sub-Image Based Features for Leaf Recognition using Support Vector Machine ", in Proceedings International Conference on Communication, Computing & security ,Pages 343-346 , 2013.

[3] A Gopal, S. Prudhveeswar Reddy and V. Gayatri," Classification of selected medicinal plants leaf using image processing", in proceedings International Conference on Machine Vision and Image Processing (MVIP),2012.

[4] R.Janani, A.Gopal," Identification of selected medicinal plants leaf using Image Features and ANN", in proceedings International Conference on Electronic Systems, 2013.

[5] Wu, Stephen Gang, et al. "A leaf recognition algorithm for plant classification using probabilistic neural network", IEEE international symposium on signal processing and information technology, 2007.

[6] Meeta, K., K. Mrunali, P. Shubhada, P. Prajakta, and B. Neha. "Survey on techniques for plant leaf classification." International Journal of Modern Engineering Research (IJMER) vol no. 2, pp.538-544, 2012.

[7] Jamil N, Hussin NA, Nordin S, Awang K. "Automatic Plant Identification: Is Shape the Key Feature?" Procedia Computer Science, vol.74, pp.36-42, December 2015.

[8] Anami, Basavaraj S., S. Nandyal Suvarna, and A. Govardhan. "A combined color, texture and edge features based approach for identification and classification of Indian Medicinal plants." International Journal of Computer Applications, pp. 45-51, 2010.

[9] Dr.H.B.Kekre, Sudeep D.Thepade, Tanuja K. Sarode and Vashali Suryawanshi, "Image Retrieval using Texture Features extracted from GLCM, LGB and KPE", International Journal of Computer Theory and Engineering, Vol. 2, No. 5, pp. 1793-8201, October, 2010.

[10] E. Sandeep Kumar, Viswanath Talasila, "Recognition of Medicinal Plants based on its Leaf Features", Froc. Of National Systems Conference, llT- Jodhpur, India, December 2013.

[11] D. Venkataraman, Narasimhan, S., Shankar, N., S. Sidharth, V., and D. Prasath, H., "Leaf recognition algorithm for retrieving medicinal information", Advances in Intelligent Systems and Computing, vol. 530, pp. 177-191, 2016..