Recognition of Ayurvedic Medicinal Plants from Leaves: A Computer Vision Approach

Amala Sabu
Department of Computer Science
College of Engineering Poonjar
Kerala,India
amalasabu1992@gmail.com

Sreekumar K
Department of Computer Science
College of Engineering Poonjar
Kerala, India
sree@cep.ac.in

Rahul R Nair
Robotics &Cognitive Systems Group
Tata Consultancy Services, Kochi
Kerala, India
rahul7manu@gmail.com

Abstract—Plants are an indispensable part of our ecosystem and India has a long history of using plants as a source of medicines. Since the advent of modern allopathic medicine, the use of traditional medicine declined to a considerable extent. However, in recent years, traditional medicine has made a comeback for a variety of reasons like they are inexpensive, nontoxic and does not impact any side effect. Different kind of medicinal plant species are available on earth but it is very difficult to identify the plant. Considerable knowledge accumulated by the villagers and tribal on medicine from plants remains unknown to the scientists and urban people. This kind of knowledge is usually handed down through generations. Our immediate concern is to preserve this knowledge in digital form through the concepts of machine learning, pattern recognition and computer vision. A machine can identify a medicinal plant through the features extracted from the leaf images, together with a classification algorithm. This paper proposes a computer vision approach for the recognition of ayurvedic medicinal plant species found in Western Ghats of India. The proposed system uses a combination of SURF and HOG features extracted from leaf images and a classification using k-NN classifier. Our experiments show results which seem to be sufficient for building apps for real life use.

Keywords –Plant Recognition; Leaf Recognition; HOG; SURF; Computer Vision; Machine Learning;

I. Introduction

The living beings on the earth are dependent upon the plants. The plants are the main source of oxygen. Apart from this, the plants are used in a variety of industrial applications such as herbs, ingredients in biofuels, biomass, pharmaceuticals etc. People have been using plant as a traditional medicine. It's considered non-expensive and does not cause any side effects. Different kinds of plant species are available on the earth; recognition of these plants is difficult. There are several methods to recognize a plant. At present, plants are identified manually by taxonomist, which are prone to human errors. To avoid this, automated plant identification system has been developed by several researchers. Medicinal plants are those plants that are used in treating and preventing specific ailments and diseases that affect human beings. Knowledge accumulated by the villagers and tribal on herbal medicine remains unknown to the scientists and urban people. Many plant species associated with the rural people are on the verge of disappearing and are on vulnerable list. The impact of deforestation, urbanization and modernization is shifting the rural people from their natural habitats and their very knowledge particularly with respect to herbal drugs is slowly disappearing. Our immediate goal is to preserve this knowledge. Whatever knowledge exists today is mostly confined to older generation. In this context some approaches needed for the preservation and development of traditional knowledge. This paper presents an automated system to recognize species of plants by analyzing the digital images of their leaf. For this recognition, we are using traditional knowledge from old generation. Several methodologies have been proposed to analyze plant leaves in an automated fashion

II. RELATED WORKS

Researchers have tried many methodologies to extract the features and identify the plant species automatically. Most of these methods make use of the combination of many parameters like color, shape, texture features etc. In [1], the statistical parameters like arithmetic mean, standard deviation, Convex Hull Ratio, eccentricity and entropy of the plant leaves images are considered for the classification of plant leaves. In[2], propose a leaf recognition system for plant species classification using leaf image data through a novel direct acyclic graph based multi-class least squares twin support vector machine (DAG-MLSTSVM) classifier. Hybrid feature selection (HFS) approach is used to obtain the best discriminant features for the recognition of individual plant species. Leaves are recognized on the basis of shape and texture features. In[3], the system works by extracting features like length, breadth, aspect ratio, diameter, leaf perimeter and rectangularity from leaf and used the concept of decision trees to train the dataset and classify the leaves. In paper[5], we propose a new approach for plant leaf classification, which treat histogram of oriented gradients (HOG) as a new representation of shape, and use the Maximum Margin Criterion (MMC) for dimensionality reduction. In[6],proposes a leaf based plant identification method using SURF features in combination with Bag of Words and supervised learning. In[7], proposes an innovative methodology for the recognition of plant species by using a combination of shape and texture features from leaf images. The leaf shape is modeled using Curvelet Coefficients and Invariant Moments while texture is

978-1-5090-6734-3/17/$31.00 ©2017 IEEE

modeled using a Ridge Filter and some statistical measures derived from the filtered image. To classify images to predefined classes, a Neuro fuzzy classifier is used. In[10], feature extraction phase derives the common Digital Morphological Features(DMF) from fundamental features. The main contribution of this approach is the Support Vector Machine (SVM)classification for efficient leaf recognition. The paper[15],presents two advanced methods for comparative study in the field of computer vision. The first method involves the implementation of the Scalar Invariant Fourier Transform (SIFT) algorithm for the leaf recognition based on the key descriptors value. The second method involves the contour-based corner detection and classification which is done with the help of Mean Projection algorithm.

III. PROPOSED SYSTEM

The proposed system contains the four stages, namely, Image acquisition, Image pre-processing, Feature Extraction and Classification. The system follows supervised machine learning and it is completed in two phases, training and testing.

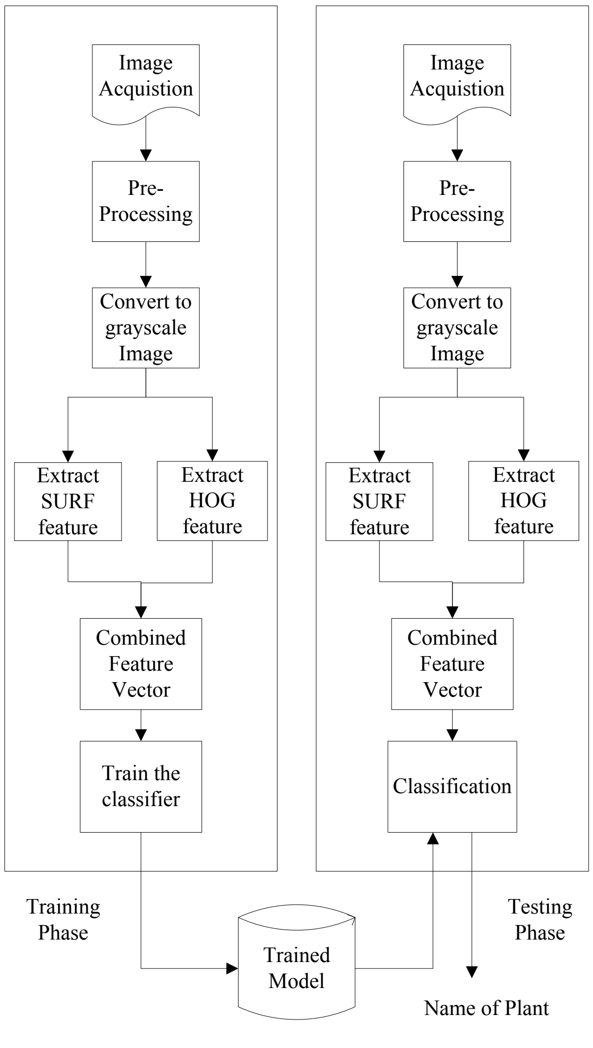

Figure 1: Proposed System Architecture

A. Image Acquisition

Leaf images used in this work were captured using a digital camera of 14.1 Megapixels in natural light with camera position vertically over the leaf. Background may be present in the images, which are eliminated later in the pre-processing step.

Figure2: Sample Images Captured [a b c d]
(a)Azadirachta indica (b) Plectranthus amboinicus (c) Justicia adhatoda (d) Ocimum tenuiflorum

B. Image Pre-processing

Pre-processing is primarily used here for removing the background area which may otherwise dominate in the image to be analyzed. First, we identified a maximum area bounding box in each image which may correspond to the leaf area. Then, this area is cropped out and taken as leaf image. Images were transformed into binary and morphological operations were performed in the pre-processing step. All the cropped images were then resized into uniform height of 1200 pixels without destroying the leaf aspect ratio.

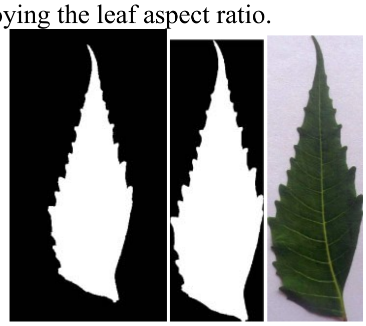

Figure3: Image Pre-processing [a b c]
(a)Background dominated input image in binary (b)Image enclosed in perfect-fitting bounding box (c) Corresponding cropped-out color image.

C. Feature Extraction

Image features we considered in this work were Speed Up Robust Feature (SURF) and Histogram of Oriented Gradients (HoG). We aim to get the scale invariance property, which is a required property as leaf images are captured in different scales in practical situations, from SURF and the unique identification property is expected to obtain from HoG feature as the leaves are having large number of veins, which could be accurately modeled using oriented gradients.

SURF feature descriptor is computed from a set of interest points which are first identified from the image. As the number of interest points may be large and since it takes much time to extract the descriptors, we extracted the descriptors from the strongest 20 interest points only. This leads to a SURF feature vector of dimensionality1280 (20*64).

An oriented gradient is present at every point in an image and so the dimensionality of HoG feature changes with image resolution. We limited the size of HoG feature vector to 1,00,000 values maximum.
These feature vectors are then fused to get a single normalized feature vector of size 1,01,280.

i. SURF

From the study of previous works in the area it is seen than some features work well for some classes and some for the other. SURF is one which is invariant of the affine transformations like scaling and rotation. It is also fast to compute. The extraction of SURF feature is based on the concept of integral image, which is calculated using the following equation.

$$S(x,y) = \sum_{i=0}^{x}\sum_{j=0}^{y} I(x,y) \qquad (1)$$

To detect interest points, SURF uses an integer approximation of the determinant of the Hessian blob detector, which can be computed with three integer operations using a pre-computed integral image. The determinant of Hessian matrix is used as a measure of local change around the point and points are chosen where this determinant is maximal. Given a point p(x,y) in an image I, the Hessian matrix H is computed using the following equation.

$$H(p,\sigma) = \begin{pmatrix} I_{xx}(p,\sigma) & I_{xy}(p,\sigma) \\ I_{yx}(p,\sigma) & I_{yy}(p,\sigma) \end{pmatrix} \qquad (2)$$

The elements are the convolution of the second order derivative of Gaussian with the image I(x,y)at the point x. For making the feature scale invariant, the image is first scaled with all possible scaling factors and in each scaled image, the high intensity points are calculated separately. Then we get a number interest points for an image. From these interest points, a strongest point selection procedure is applied in which those below a threshold are discarded. Next step is to describe these points. A 64-dimensional feature vector is extracted for each of the detected key points, which together forms a feature matrix. Next, we need to convert the feature matrix to a feature vector. In this work 20 strongest points are selected. Thus, we get a feature vector of SURF with 1280 elements.

ii. HOG

Histogram of Oriented Gradients is a feature descriptor used in computer vision and image processing for the purpose of object detection. This technique counts the occurrences of gradient orientation in localized portion of an image.HOG computed over a dense grid of cells, which are uniformly spaced, and for improving accuracy, uses an overlapping local contrast normalization procedure. The first step of calculation is the computation of the gradient values. The most common method is 1-D centered, point discrete derivative mask in one or both of the horizontal and vertical directions. This method requires the following kernels.

$$[-1\ 0\ 1] \text{ and } [-1\ 0\ 1]^T \qquad (3)$$

In the second step, cell histograms are created and each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values found in the gradient computation. These cells are then grouped together into a larger spatially connected block. This grouping is required to locally normalize the gradient strengths, in order to account for changes in illumination and contrast. The HOG feature vector is then extracted as the vector obtained from components of the normalized cell histograms from all of the block regions. These blocks may overlap, such that, each cell contributes to more than once to the final descriptor.

D. Classification

We used the simplest supervised classification algorithm 'K-Nearest Neighbor' for classifying the leaves. The dataset was partitioned into two equal parts and trained the k-nn classifier using one half and remaining half was used for testing. The inputs supplied to the k-nn classifier were the fused feature vectors and corresponding class labels. We used the Euclidean distance for measuring distance between feature points in the vector space.

The K-NN classifier is given some data points for training and also a new unlabeled data for testing. The aim is to find the class label for the new point. The algorithm has different behavior based on k,where k is the number of neighbors considered. k-nearest neighbor algorithm is a method for classifying objects based on closest training examples in the feature space. Training process for this algorithm only consists of storing feature vectors and labels of the training images. In the classification process, the unlabeled query point is simply assigned to the label of its k nearest neighbors. Typically, the object is classified based on the labels of its k nearest neighbors by majority vote. If k=1, the object is simply classified as the class of the object nearest to it.

IV. RESULTS AND ANALYSIS

A. Experimental Setup

We created our own dataset for experiments since a standard data set is not available in the domain. The images of leaves were captured using a digital camera of resolution 14.1 MP. The data set contains a total of 200 images spread across 20 classes. Each class contains images of 10 distinct leaves collected from different plants. The experiments used a two-fold cross validation in which 5 images per class were used for training and remaining 5 were used for testing the model, and thereafter the roles were reversed and average results were taken. The selection of images into testing and training sets was performed at random.

The proposed system was evaluated for performance using statistical approaches by computing the metrics Precision, Recall, F-measure and Accuracy. Table I shows the name of the medicinal plants in Malayalam Language, corresponding scientific name and the notation followed by us for easy representation in experiments. Table II shows the confusion matrix obtained in classification. Finally, Table III shows the overall effectiveness of the recognition system indicating the accuracy obtained for each class.

The overall Accuracy, Precision, Recall, and F measure were computed as,

Precision = TP/TP+FP
Recall = TP/TP+FN
Accuracy = TP+TN/TP+TN+FP+FN
F measure = 2(Precision*Recall/Precision + Recall)

Where TP is True Positive Rate, FP is False Positive Rate, TN is True Negative Rate and FN is False Negative Rate.

B. Discussion

The main motive behind this work was to develop a plant species classification approach, which doesn't rely on the peripheral characteristics of leaves. So we didn't consider the modeling of leaves using parameters like area, perimeter, aspect ratio, curvature etc., which might follow a standard probability distribution like Gaussian.

Since we focused on the target domain of 'Ayurvedic medicinal plants of the Western Ghats of India', the standardization of results was difficult since there were no benchmark for results which used identical leaves as we used in our experiments. So our results couldn't be compared against other plant species recognition works. The dataset we created still represents a very narrow band of the whole medicinal plant species spectrum available in the Western Ghats.

The main highlight or novelty we claim is the use of two features which are independent of the shape and morphology of leaves. SURF was used in-order to achieve the scale and rotation invariance whereas the role of HoG was to discriminate the leaf categories by considering the orientation of the veins and small netted veins. From our experiments, it has been revealed that HoG can efficiently capture the surface texture composed of veins and netted sub-veins so as to give high accuracy. The computational time also seemed to be low, which is suitable for real-time recognition using hand held devices too.
More experiments can be done to see the trade-off between SURF and HoG features in recognition, to see the effect in recognition with reduction in HoG dimensionality, and also to see the impact in result with increase in the no. of feature points of SURF.

V. CONCLUSION

This paper has proposed and implemented a system for automatic identification of medicinal plants from their leaves. The proposed system makes use of computer vision and machine learning approaches to identify a pre-trained medicinal plant from its leaf. Main highlight of this work is the non-use of typical shape and color features of leaves which are computationally expensive to extract as they are spatial features. The proposed system uses a blend of SURF and HOG features which provided nearly 100% of accuracy when experimented with k-nn classifier. Scope for future work includes expanding the leaf dataset, changing the classifier from k-nn to SVM or ANN and to experiment with a blend of more features added with HOG and SURF.

Table I: Naming Convention

Medicinal Plant Name	Scientific Name	Notation
Aaryaveppu	Azadirachta indica	A
Curryveppu	Murraya koenigii	B
Karintakara	Costus Igneus	C
Mehandhi	Lawsonia inermis	D
Cheera	Amaranthus	E
Panel	Uvaria narum	F
Panikoorkka	Plectranthus amboinicus	G
Murikootti(type 1)	Hemigraphis colorata	H
Muringa	Moringa oleifera	I
Sarppagandhi	Rauwolfia serpentina	J
Narakam	Citrus medica	K
Sarvasugandhi	Pimenta dioica	L
Kodangal	Centella asiatica	M
Chembarathi	Hibiscus rosa-sinensis	N
Thumba	Leucas zeylanica	O
Pera	Pyrus	P
Aadalodagam	Justicia adhatoda	Q
Kuruthotty	Aadalodagam	R
Murikootti(type 2)	Green flame ivy	S
Thulasi	Ocimum tenuiflorum	T

Table II: CONFUSION MATRIX
Predicted Class

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	1
I	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	1	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	4	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5

Table III: PERFORMANCE MATRIX

Class Name (in notation)	TP	FP	FN	TN	Precision	Recall	Accu-racy	F-measure
A	5	0	0	95	1	1	1	1
B	5	0	0	95	1	1	1	1
C	5	0	0	95	1	1	1	1
D	5	0	0	95	1	1	1	1
E	5	0	0	95	1	1	1	1
F	5	0	0	95	1	1	1	1
G	5	0	0	95	1	1	1	1
H	4	0	1	95	1	0.8	0.99	0.89
I	5	0	0	95	1	1	1	1
J	5	0	0	95	1	1	1	1
K	4	0	1	95	1	0.8	0.99	0.89
L	4	2	1	93	0.67	0.8	0.97	0.73
M	5	0	0	95	1	1	1	1
N	5	0	0	95	1	1	1	1
O	5	0	0	95	1	1	1	1
P	4	1	1	94	0.8	0.8	0.98	0.8
Q	5	0	0	95	1	1	1	1
R	5	0	0	95	1	1	1	1
S	5	0	0	95	1	1	1	1
T	5	1	0	94	0.83	1	0.99	0.91

From the above performance matrix, it is revealed that,

Average accuracy = 0.996, Average F-measure = 0.961, Average Precision = 0.965, and Average Recall = 0.96

References

[1] Pushpa BR, Anand C and Mithun Nambiar P ”Ayurvedic Plant Species Recognition using Statistical Parameters on Leaf Images” International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 7 (2016)

[2] Divya Tomar and Sonali Agarwal ”Leaf Recognition for Plant Classification Using Direct Acyclic Graph Based Multi-Class Least Squares Twin Support Vector Machine” Journal of Image and Graphics Vol. 16, No. 3 (2016).

[3] D Venkataraman, Siddharth Narasimhan, Shankar N, S Varun Sidharth and Hari Prasath D ”Leaf Recognition Algorithm for Retrieving Medicinal Information” Springer International Publishing AG 2016.

[4] Narasimhan,S. Prudhveeswar Reddy and V. Gayatri ”Classification of Selected Medicinal Plants Leaf Using Image Processing” IEEE 2012.

[5] Xue-Yang Xiao, Rongxiang Hu,Shan-Wen Zhang, and Xiao-FengWang”HOG-Based Approach for Leaf Classification” Springer International Publishing 2010.

[6] Quang-Khue Nguyen,Thi Lan Le and Ngoc hai Pam” leaf Based Plant Identification System for Anusing SURF Features in combination with Bag of Words Model and supervised learning” International Conference Advanced Technologies for communications,2013

[7] Jyotismita Chaki, Ranjan Parekh and Samar Bhattacharya ”Plant Leaf Recognition Using Ridge Filter and Curvelet Transform with Neuro Fuzzy Classifier” Springer India 2016

[8] Snehal Shejwal, Prachi Nikale,Ashwini Datir,Ashvini Kadus,Pratik Bhade and Rahul Pawar ”Automatic Plant Leaf Classification on Mobile Field Guide” IJCST Vol. 6, Iss ue 2, April - June 2015

[9] Shreepad S. Sawant and Preeti S. Topannavar ”Introduction to Probabilistic Neural Network Used For Image Classifications” International Journal of Advanced Research in Computer Science and Software Engineering, 2015, Volume 5

[10] ArunPriya C, Balasaravanan T and Antony Selvadoss Thanamani ” An Efficient Leaf Recognition Algorithm for Plant Classification Using Support Vector Machine” Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, March 2012

[11] R Janani and A Gopal ”Identification of selected Medicinal Plant Leaves Using Image Fatures and ANN” International Conference on Advanced Electronic Systems, 2013

[12] Sreekumar K,Anjusha B,Rahul R Nair ”Automatic Image Annotation Using Synthesis of Complementary Features” Proceedings of the 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013)

[13] Sandeep Kumar.E ”Leaf color, area and edge features based approach for identification of Indian medicinal plants” - Indian Journal of Computer Science and Engineering (IJCSE).

[14] Qunang-Khue Ngueyn ”Leaf based plant identification system for antroid using SURF features in combination with bag of words model and supervised learning” - International Conference on Advanced Technologies for communicatuons,2013

[15] Shubham Lavania ”Leaf Recognition using Contour based Edge Detection and SIFT Algorithm” - IEEE International Conference,2014

[16] Boran Sekeroglua,Yucel Inanb ”Leaves recognition system using a neural network” - 12th International Conference on Application of Fuzzy Systems and Soft Computing,2016