Real-Time Identification of Medicinal Plants using Machine Learning Techniques

Conference Paper · February 2019

DOI: 10.1109/ICCIDS.2019.8862126

5 authors, including:

Lekshmi Kalinathan
Sri Sivasubramaniya Nadar College of Engineering

Ruba Soundar Kathavarayan
Mepco Schlenk Engineering College

Jegadish KUMAR K J
Sri Sivasubramaniya Nadar College of Engineering

Real-Time Identification of Medicinal Plants using Machine Learning Techniques

[1]Sivaranjani.C
Department of Computer Science and Engineering,
SSN College of Engineering,
Kalavakkam, India
sivaranjani13c@gmail.com

[2]Lekshmi Kalinathan
Department of Computer Science and Engineering,
SSN College of Engineering,
Kalavakkam, India
lekshmik@ssn.edu.in

[3]Amutha.R
Department of Electronics and Communication Engineering,
SSN College of Engineering,
Kalavakkam, India
amutharvaag@gmail.com

[4]Ruba Soundar Kathavarayan
Department of Computer Science and Engineering,
P.S.R. Engineering College,
Sivakasi, India
rubasoundar@yahoo.com

[5]Jegadish Kumar.K.J
Department of Electronics and Communication Engineering,
SSN College of Engineering,
Kalavakkam, India
jegadishkj@ssn.edu.in

***Abstract*— The lighting condition of the environment are uncontrolled, so the segmentation of a leaf from the background is considered as a complex task. Here we propose a system which can identify the plant species based on the input leaf sample. An improved vegetation index, ExG-ExR is used to obtain more vegetative information from the images. The reason here is, it fixes a built-in zero threshold and hence there is no need to use otsu or any threshold value selected by the user. Inspite of the existence of more vegetative information in ExG with otsu method, our ExG-ExR index works well irrespective of the lighting background. Therefore, the ExG-ExR index identifies a binary plant region of interest. The original color pixel of the binary image serves as the mask which isolates leaves as sub-images. The plant species are classified by the color and texture features on each extracted leaf using Logistic Regression classifier with the accuracy of 93.3%.**

Keywords* — *ExG-ExR, Logistic Regression

I. INTRODUCTION

Medicinal Plants are fundamental to the equalization of nature and in individuals lives. They are the source of medicine. Thus, the identification of the medicinal plants is important because medicinal plants are the fundamental part of life on the Earth and allows humans and other organisms to exist by generating the oxygen and food. An advanced plant identification system can be used for quick characterization of medicinal plant species without requiring the expert advice, thus automizing their task.

Identification of plants based on the flowers and fruits needs morphological features. Those features are the number of stamens in flower and number of ovaries in fruits. Distinguishing these plants by using the keys is an extremely tedious process and are done only by the botanists. In addition to this time intensive task, the few disadvantages in distinguishing plants utilizing these features are the inaccessibility of the needed morphological data and only the experts can understand those botanical terms. The leaves play a vital role in identification of plant species. Also, the leaves can be easily collected everywhere at all seasons, whereas flowers can only be collected during blooming season.

Classification of the plant species based on the leaf requires some preprocessing. The first step is segmenting the leaf from the background where the lighting condition is uncontrolled. The next step is extracting the features based on color and texture from the segmented leaf. The third and the last step is to categorize the plant species with the above said extracted features.

II. RELATED WORK

This paper is inspired by the research work done by George E Meyer, Joao camargo Neto [1] based on the vegetative index, ExG-ExR. This provides enough support for the segmentation by providing an optimized index which is inferring from multiple channel information unlike the method discussed in [7], uses multiple thresholds for segmentation.

Liu, Albert and Yangming Huang [2] developed a plant identification system using CNN to gather the bottelneck features, called CNN codes. Finally, these CNN codes were trained with SVM for classification purpose. But this method holds for clean images. These images are characterized with leaves that are well aligned on a contrast background, with few or no variations of color or luminance.

Kumar, P. M., Surya, C. M. and Gopi, V. P. [3] used different plant features such as color, texture, shape. But this method works well only for static background or plain background.

Putzu, L., Di Ruberto, C., Fenu, G. [4] make use of saliency maps methods such as Graph-Based Manifold Ranking (GMR), Visual Saliency Feature (VSF), Gaussian Pyramids [6], based leaf extraction. But This method works well in the presence of an untextured background.

Anantrasirichai, Nantheera, Sion, L., Hannuna and Cedric Nishan Canagarajah. [5] based on marker-controlled watershed segmentation. This method still misses some areas because of reflection, shadow and disease on the leaves.

III. PROCEDURE

A. Dataset

The Dataset used for this work contains 5 classes of medicinal plant where 20 images in each class. 70% of the images is used as train data and 30% of the images is used as test data. Figure 1 below shows the sample images in the dataset.

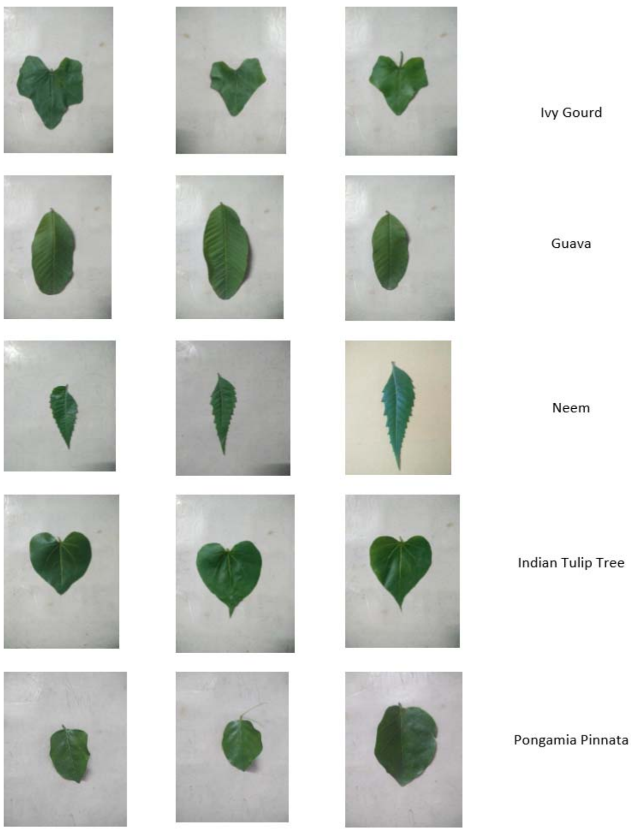

Figure-1: Sample images in the dataset

B. Proposed Work

The overview of the architecture is shown in the Figure 2. The input image is in RGB color space. The leaf in the input image is segmented using ExG-ExR method. The texture and color features can be extracted from the segmented leaf. The Weka is used to classify the medicinal plant species

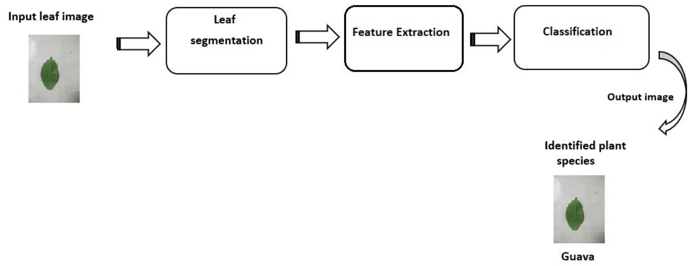

Figure-2: Overview of Architecture for the classification of plant species.

1) IMAGE SEGMENTATION

The first step is to extract the RGB channel information from the input image to detect the foreground plant information. To do that, Excess Green (ExG) channel and Excess Red (ExR) channel information are extracted as given in the Equations (1) and (2). Later, the excess red information is eliminated from the excess green information to retain more green channel information. When binary threshold is applied on the improved vegetative index, ExG-ExR information, a binary plant region of interest is accurately segmented as shown in the Figure 3. The second step is to use the original color pixel of the binary image as the mask which isolates leaves as sub-images. The segmentation of sample leaf images in the dataset using the aforementioned vegetative index is shown in the Figure 4.

$$ExG = 2*G-R-B \quad (1)$$

$$ExR=1*4*R-G \quad (2)$$

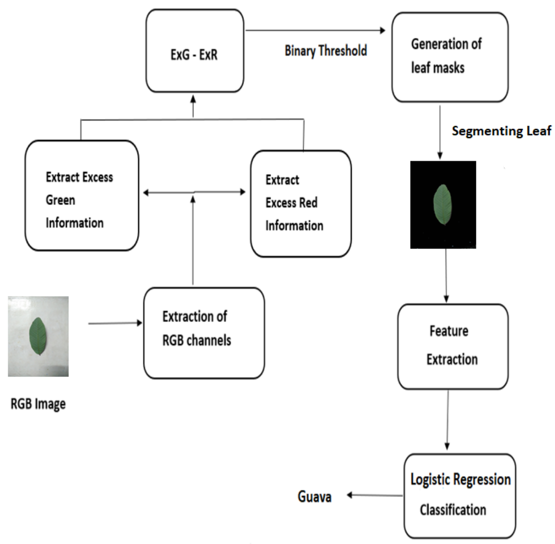

Figure-3: Detailed Architecture of the Proposed Method

978-1-5386-9471-8/19/$31.00 ©2019 IEEE

2) FEATURE EXTRACTION

- **Color Features**

The important feature in the representation of an image is color. Color moments are obtained through the statistical features such as mean and standard deviation estimated on the three channels RGB (red, green, blue). From the standard deviation and mean, the first order statistics of an image is obtained. These are based on the properties of individual pixels as

$$Mean(\mu) = \sum_{i=0}^{M}\sum_{j=0}^{N}\frac{C(i,j)}{MN} \tag{3}$$

Where:
M, N - are image dimensions
C - is one of the channels in a given RGB image
C(i,j) – color value on column i and row j

$$Standard\ Deviation(\sigma) = \sqrt{\sum_{i=0}^{M}\sum_{j=0}^{N}\frac{(c(i,j)-\mu)^2}{MN}} \tag{4}$$

- **Texture Features**

The texture of the image can be measured by the co-occurrence matrix, various dimensions of color or the grayscale values of the image. Features that are generated using this technique is known as Haralick features. Haralick feature such as Contrast, Correlation, Inverse difference moments, Entropy are calculated.

i) Contrast: Contrast reflects the sensitivity of the textures in relation to changes in the intensity an aslo measures the quantity of local changes in an image. The measure of intensity of contrast between a pixel and its neighborhood is returned.

$$Contrast = \sum_{n=0}^{G-1}\{\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}P(i,j)\} \tag{5}$$

where *G* represents the no. of gray levels used, and *P* represents probability distribution of GLCM.

ii) Correlation: Correlation measures the gray tone linear dependencies in the image. The range of the feature values are from -1 to 1, the extremes indicates perfect negative and positive correlation respectively.

$$correlation = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu_i)P_d(i,j)}{\sigma_i\sigma_j} \tag{6}$$

where $\mu_i, \sigma_i, \sigma_j$ - are the mean and the standard deviation of *P*.

iii) Inverse Difference Moment (IDM): It is usually called homogeneity that measures the local homogeneity of an image and also measures the image texture. The measure of the closeness of the distribution of the GLCM elements to the GLCM diagonal is obtained by IDM feature. IDM has a range of values which determines textured or non-textured image.

$$Inverse\ difference\ moment = \sum_i\sum_j\frac{1}{1+(i-j)^2}P(i,j) \tag{7}$$

iv) Entropy: It is a measure of randomness of intensity image

.

$$Entropy = -\sum_i\sum_j P(i,j)\log(p(i,j)) \tag{8}$$

3) CLASSIFICATION

Logistic Regression classifier is used for identifying the medicinal plant. Since, our proposed method falls on the classification of multiple classes, the binary logistic regression model has been extended to multinomial logistic regression. Through a combination of binary logistic regressions, multiple groups are compared by the multinomial logit . This allows each category of the label obtained to be compared to a reference category in the dataset. Typically, the category with the most noteworthy score is selected as the reference category. On evaluating the features of the images in our dataset on two different multiple classes classifiers shown in Table 2, logistic regression outperforms the multi-class classifier with the accuracy of 93.3%. The results of classification is done by using Weka.

C. Experimental Results

Figure-4: Examples of medicinal plants (a) Original images, (b) ExG images, (c) ExR images and (d) Improved vegetation index (ExG-ExR) images.

Table-1: Statistical and Texture features of sample leaves in the dataset

Features	**Species**				
	IG[b]	**Guava**	**IT**[c]	**PP**[d]	**Neem**
Mean_r	12.9	4.0	7.7	3.1	4.9
Mean_g	18.3	6.8	12.1	5.0	7.1
Mean_b	11.7	4.4	7.5	3.3	4.3
Stddev_r	26.9	15.4	19.7	13.7	17.7
Stddev_g	37.3	26.1	29.4	21.7	25.3
Stddev_b	24.4	17.1	19.0	14.6	15.6
Contrast	138.4	90.7	102.6	48.1	69.8
Correlation	0.9	0.9	0.9	0.9	0.9
IDM[a]	0.8	0.9	0.9	1.0	0.9
Entropy	2.9	2.3	2.3	0.8	1.2

[a] = Inverse_difference_moments, [b] = Ivy Gourd, [c] = Indian Tulip, [d] = Pongamia Pinnata

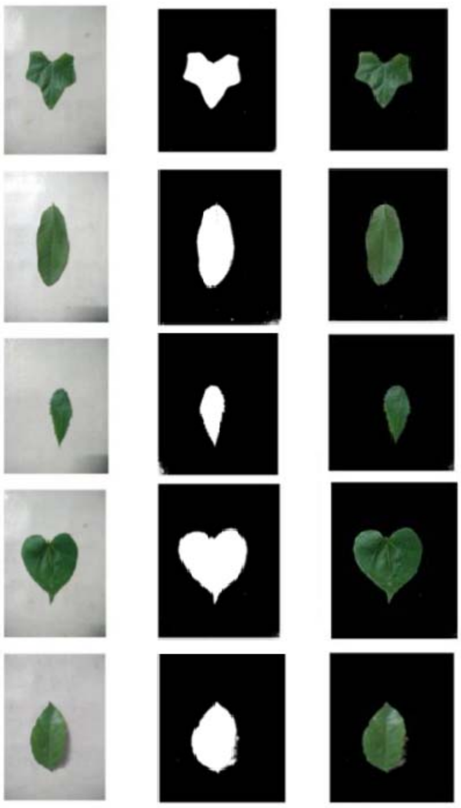

Figure-5: Automated leaf segmentation using binary masks obtained from ExG-ExR index.

Table-2: Performance Comparison of two different multiple classes classifiers

Classifier	**Accuracy**
Logistic Regression	93.3%
Multi-class classifier	80%

IV. CONCLUSIONS

In this work, we addressed the problem of identifying the medicinal plant species by the analysis of leaf images obtained directly from their habitat and irrespective of lighting conditions. The fixed zero threshold, ExG-ExR vegetative index is successfully tested for image dataset. The result shows that the algorithm can adequately segment the leaf region. This method worked well in images with reflection. The feature extraction based on the color and texture features is done. The classification of medicinal plant species is done by using Weka and the accuracy of 93.3% is measured. In future we have planned to design and develop a system which automatically identifies plant species through the analysis of not only the leaf images also the other parts of the plant acquired directly in their habitat irrespective of complex backgrounds and various lighting condition.

REFERENCES

[1] G. E. Meyer and J. C. Neto, "Verification of color vegetation indices for automated crop imaging applications," Computers and Electronics in Agriculture, vol. 63, October 2008, pp. 282-293.

[2] Liu, Albert and Y. Huang. "Plant Leaf Recognition." conference proceedings, 2016.

[3] P. M. Kumar, C. M. Surya, and V. P. Gopi, " Identification of ayurvedic medicinal plants by image processing of leaf samples," 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, 2017, pp. 231-238.

[4] L. Putzu, C. Di Ruberto and G. Fenu, "A Mobile Application for Leaf Detection in Complex Background Using Saliency Maps," In: Blanc-Talon J., Distante C., Philips W., Popescu D., Scheunders P. (eds) Advanced Concepts for Intelligent Vision Systems. ACIVS 2016. Lecture Notes in Computer Science, 2016, vol. 10016, Springer, Cham.

[5] Anantrasirichai, Nantheera, L. Sion, Hannuna and C. N. Canagarajah. "Automatic Leaf Extraction from Outdoor Images," CoRR abs/1709.06437, 2012.

[6] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit and C. H. Anderson, "Overcomplete steerable pyramid filters and rotation invariance," 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 1994, pp. 222-228.

[7] K. Lekshmi, K. Ruba Soundar, E. Abinaya, P. Gayathri and T. Keerthana, "Automated selection of parameters using Tabu Search in image segmentation", 2016 10th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, 2016, pp. 1-4.