ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/282148402

An automatic leaf recognition system for plant identification using machine vision technology

Article · April 2013

CITATIONS
28

READS
299

3 authors, including:

Shanu Sharma
Amity University
7 PUBLICATIONS **58** CITATIONS
SEE PROFILE

All content following this page was uploaded by Shanu Sharma on 25 September 2015.

The user has requested enhancement of the downloaded file.

AN AUTOMATIC LEAF RECOGNITION SYSTEM FOR PLANT IDENTIFICATION USING MACHINE VISION TECHNOLOGY

VIJAY SATTI*
Student, CSE Department, ASET, Amity University
Noida, Uttar Pradesh, India
vijay.satti@live.com

ANSHUL SATYA
Student, CSE Department, ASET, Amity University
Noida, Uttar Pradesh, India,
anshulksatya@yahoo.com

SHANU SHARMA
Assistant Professor, CSE Department, ASET, Amity University
Noida, Uttar Pradesh, India
shanu.sharma1611@gmail.com

Abstract:

Plants are the backbone of all life on Earth and an essential resource for human well-being. Plant recognition is very important in agriculture for the management of plant species whereas botanists can use this application for medicinal purposes. Leaf of different plants have different characteristics which can be used to classify them. This paper presents a simple and computationally efficient method for plant identification using digital image processing and machine vision technology. The proposed approach consists of three phases: pre-processing, feature extraction and classification. Pre- processing is the technique of enhancing data images prior to computational processing. The feature extraction phase derives features based on color and shape of the leaf image. These features are used as inputs to the classifier for efficient classification and the results were tested and compared using Artificial Neural Network (ANN) and Euclidean (KNN) classifier. The network was trained with 1907 sample leaves of 33 different plant species taken form Flavia dataset. The proposed approach is 93.3 percent accurate using ANN classifier and the comparison of classifiers shows that ANN takes less average time for execution than Euclidean distance method.

Keywords: plant Identification; features extraction; neural network; Euclidean distance.

1. Introduction

Plants are essential to the balance of nature and in people's lives. They are the ultimate source of food and metabolic energy for nearly all animals, which cannot manufacture their own food. Thus the study of plants is vital because they are a fundamental part of life on Earth, and generate the oxygen and food that allow humans and other organisms to exist. A digital plant identification system can be used for quick characterization of plant species without requiring the expertise of botanists, thus automizing their task.

This paper describes our approach for the plant identification using digital images of leaves. Leaf-based features are preferred over fruits, flowers, root, stem etc. due to the seasonal nature of the fruits & flowers and inequality in root & stem characteristics. There are different publicly available leaf image datasets such as Flavia dataset, Leafsnap dataset, Intelengine dataset, ImageCLEF dataset and many others. The performance of this experiment is evaluated using Flavia dataset.

2. Literature Survey

Although a significant amount of research has been done studying various aspects of leaf identification in inventory systems, most of it deals with semi-automated systems. A state-of-the-art system which is fully automated and requires least human interaction is yet to be developed.

Arora. A et al. [1] categorized the different images and used a variety of novel pre-processing methods such as shadow and background correction, petiole removal and automatic leaflet segmentation for identifying the leaf blobs. Also used complex network framework along with novel tooth detection method and morphological operations to compute several useful features. They used the Pl@ntLeaves II dataset.

Pavan et al. [3] proposed an algorithm for identification using multiclass classification based on color, shape volume and cell feature. They performed three stage comparisons: first stage compares redness, greenness, blueness index feature, second stage compare shape feature and the last stage compares cell feature and volume fraction feature. Experiment is performed on a sample of diverse collection of 1000 leaf and flower images. Limitations of this approach is that it semi-automatic approach and its recognition rate is up to 85% percent on an average.

Arun Priya [4] the proposed approach consists of three phases such as preprocessing: transforming to gray scale and boundary enhancement, feature extraction: derives the common DMF from five fundamental features and classification: Support Vector Machine (SVM) classification for efficient leaf recognition. 12 leaf features which are extracted and orthogonalized into 5 principal variables are given as input vector to the SVM.

Valliamal et al. [5] A probabilistic curve evolution method with particle filters is used to measure thesimilarity between shapes during matching process. The experimental results prove that the preferential image segmentation can be successfully applied in leaf recognition and segmentation from a plant image.

Dr. H.B.Kekre et al. [6] the method of CBIR is discussed in this paper to filter images based on their content. In this paper feature vector is generated using color averaging technique, similarity measures and performance evaluation. Precision –Recall cross over plot is used as the performance evaluation measure to check the algorithm. The effect due to the size of database and number of different classes is seen on the number of relevancy of the retrievals.

Javed et al. [7] used PNN to classify the plants with broad flat leaves. In this algorithm there were few select point where the user needs to specify the leaf blades and a base point according to which the image is then aligned and compared with other images on the basis of some features like area, eccentricity, etc. They used 1200 sample leaves belonging to 30 different plants to train their system. This system is also semi automatic and 91.41 percent accurate.

In literature survey it has been observed that most of the systems emphasized basically on morphological features only, some have used tooth features also. So, in this paper the new algorithm has been designed by extracting 5 geometric feature, 12 morphological features, tooth features and color features also to increase the efficiency of proposed leaf recognition system. The algorithm is explained in the next section.

3. Proposed Methodology

A typical image based plant identification system is shown in fig. 1 and the major steps are explained in consecutive sub-sections.

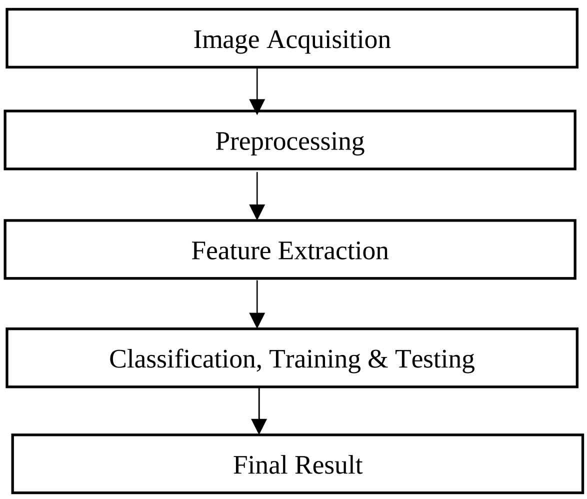


Fig. 1. Flow diagram of proposed scheme.

3.1. *Image acquisition*

A leaf image can be easily acquired using scanner or digital camera. The image can be of any size. However, for better results, the image should have preferably single color background with no petiole. The proposed system is tested on Flavia dataset which contains 1907 RGB leaf images of 33 plants; each species has 40 to 60 sample leaves. Each image in dataset is of 1600x 1200 resolution having white background and with no leafstalk. File names of all images are 4-digit numbers, followed by a ".jpg" suffix.

3.2. *Preprocessing*

In order to extract any specific information, image preprocessing steps are carried out before the actual analysis of the image data. Preprocessing refers to the initial processing of input leaf image to eliminate the noise and correct the distorted or degraded data. Fig. 2 illustrates techniques like grayscale conversion, binarization, smoothing, filtering, edge detection, etc. used for the enhancement of the leaf image.

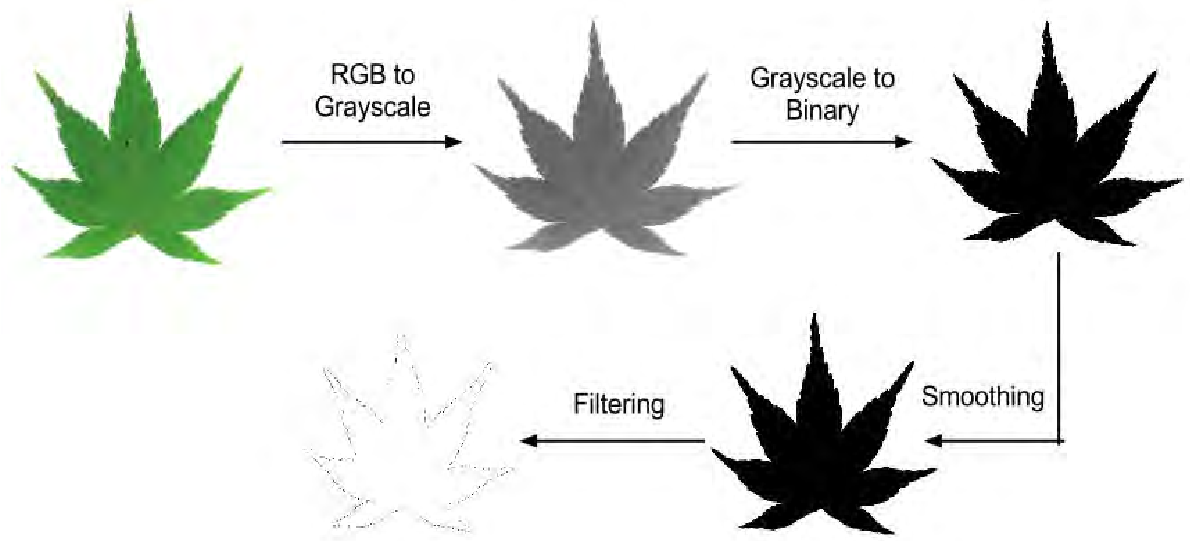


Fig. 2. Preprocessing steps performed on an *Acer Palmatum* leaf image.

3.3. *Feature extraction*

Our method takes into account the color and shape features of the leaf. Leaves of different plants are invariably similar in color and shape therefore a single feature alone may not produce expected results.

3.3.1. *Color features*

The method of image searching and retrieval proposed by Dr. H.B. Kekre et al. [7] mainly focuses on the generation of the color feature vector by calculating the average means. In the proposed algorithm, first the three color planes namely Red, Green and Blue are separated. Then for each plane row mean and column mean of colors are calculated. The average of all row means and all columns means is calculated for each plane. The features of all 3 planes are combined to form a feature vector. Once the feature vectors are generated for an image, they are stored in a feature database.

3.3.2. *Shape features*

We defined shape features on the basis of morphological features and tooth features:

A) *Geometric features*

We used the similar (as described in [8]) commonly used 5 geometric features (DMFs), illustrated in fig. 3, derived from following 5 basic features:

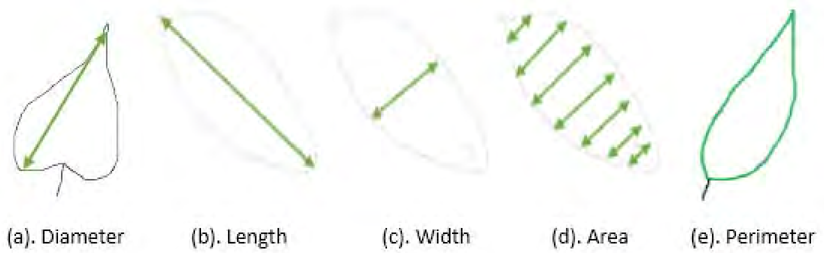


Fig. 3. The five basic morphological features.

(1). *Diameter*: The diameter of the leaf is the longest distance between any two points on the closed contour of the leaf.

(2). *Physiological Length*: It is the length of the line connecting the two terminal points of the main vein in the leaf.

(3). *Physiological Width*: It refers to the distance be-tween the two endpoints of the longest line segment perpendicular to the physiological length.

(4). *Leaf Area*: It is the number of pixels of binary value 1 on smoothed leaf image.

(5). *Leaf Perimeter*: It is the number of pixels along the closed contour of the leaf.

B) *Morphological Features*

Based on above 5 basic geometric features, we can define following 12 digital morphological features:

(1). *Smooth Factor*: This is defined as the ratio between area of leaf image smoothed by 5x5 rectangular averaging filter and the one smoothed by 2x2 rectangular averaging filter.

(2). *Aspect Ratio*: This is defined as the ratio of physiological length to physiological width, i.e., L/W.

(3). *Form Factor:* It is defined as the difference between a leaf and a circle and is calculated by the formula $4\pi A/P^2$.

(4). *Rectangularity:* It describes how similar a leaf is to a rectangle and is computed as L.W/A

(5). *Narrow Factor:* It defines the narrowness of the leaf and is calculated as D/L.

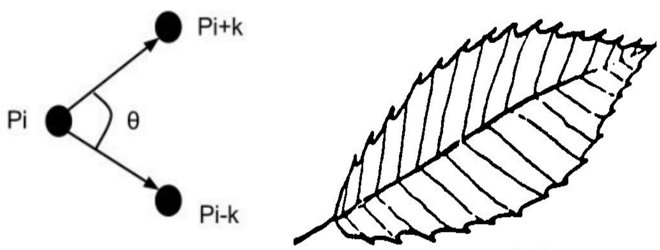
(6). *Perimeter Ratio of Diameter:* It is defined as the ratio of the perimeter of the leaf to the diameter of the leaf, i.e., P/D.

(7). *Perimeter Ratio of Physiological Length and Physiological Width:* It is defined as the ratio of the perimeter of the leaf to the sum of its physiological length and physiological width, i.e., P=(L+W).

(8). *5 Vein Features:* Leaf vein forms the basis of leaf characterization and classification as they define the skeletal structure of the leaf. Different species have different leaf vein patterns which can be used in distinguishing the leaves that have similar shape. The standard procedure for computing the vein features is to perform a morphological opening operation on the grayscale image. A flat, disk shaped structuring element of radius 1,2,3,4 is used and the resultant image is then subtracted from the contour of the leaf. The output resembles to the vein structure of the leaf on the basis of which following 5 vein features are calculated: A_1/A, A_2/A, A_3/A, A_4/A, A_4/A_1 where A_r is the remaining leaf obtained using a structuring element of radius r and A is the area of the leaf.

C) *Tooth features*

A tooth [1] in a leaf is a pixel that is serrated and toothed around the margins of the leaf. Fig. 4. (a) represents a tooth point of a leaf. To determine whether a point P_i on the margin of the leaf is a tooth point or not, we examine the angle θ subtended at P_i by its neighbors P_{i-k} and P_{i+k} (where k is the threshold). If the angle θ is within a particular range, then P_i is a tooth; otherwise, it is not. Fig. 4. (b) represents a standard toothed leaf. In proposed algorithm total number of tooth points are calculated in each image and stored against feature vector.



(a) **(b)**

Fig. 4. (a). A tooth point (b). Toothed leaf

3.4. *Classification, training & testing*

General statistical classification is the process of identifying a set of categories, or classes, to which a new observation belongs, on the basis of prior knowledge such as a training dataset. More specifically, classification in this work will be the process used to assign a certain plant species to an image, based on its feature set. It is also a subset of the more general classification problem in statistics and machine learning, namely supervised learning. We formalize the classification elements as follows:

Where C represents the class set, f a feature vector in the corresponding m-dimensional feature space F, a_i is a feature attribute and T the training set of feature vectors and their respective class labels. Hence, we are looking for a function $Class(f) : F \rightarrow C$ that assigns a class label from C to a given feature vector f based on the data from T. Throughout this work we will be looking for a classification method not only capable of attributing a class label to a feature vector f_s but also a confidence vector describing the probability that a given sample belongs to a certain class:

3.4.1. *Classifier selection*

We followed two approaches to classify our dataset, i.e., Neural Networks and Euclidean Distance Method.

A neural network, illustrated in fig. 5, consists of units (neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally the networks are defined to be feed-forward: a unit feeds its output to all the units on the next layer, but there is no feedback to the previous layer. Weightings are applied to the signals passing from one unit to another, and it is these weightings which are tuned in the training phase to adapt a neural network to the particular problem at hand. This is the learning phase.

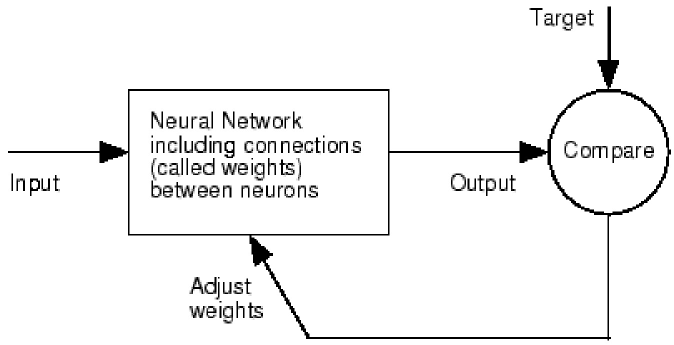


Fig. 5. A typical Neural Network

The Euclidean or KNN classifier based on the distance is direct and simple. Special interest was given to KNN due to its Simplicity and Efficiency. It is one of the simplest classifiers with characteristics fitting our requirements. It requires no training computations and is easily handled by weak processors. Its testing time, however, grows linearly with the size of the training set, limiting the scalability of the classifier. It is based on distance measures in feature space but instead of comparing f_s to a class representative value, it compares it to all samples of the training set $f_t i$ selecting the first k closest ones. We call the subset of k-closest training samples K.

3.4.2. *Training & Testing*

Flavia dataset contains a total of 1907 images of 33 different plant species. These images were used to train the classifier. For each type of plant in flavia dataset, we selected 5 species of leaves from testing sets which are then used to test the efficiency of the proposed algorithm in terms of accuracy and execution time.

4. Experimental results

The classification was performed using Neural Networks and Euclidean classifier. The results obtained with these schemes were used to compare the classification technique and to conclude this study. Table I shows the accuracy of the system for KNN & ANN classifier.

Table I. Accuracy for the KNN & ANN classifier.

Classifier	Accuracy (%)
KNN	85.9
ANN	93.3

The time taken and the accuracy attained with the classification scheme are the important viewpoints of any user. It is found that ANN classification scheme is better than KNN for a dataset with large number of images whereas, KNN outperforms the ANN approach for a smaller dataset. The efficiency is calculated not only in terms of accuracy but also in terms of time taken by the classifier used. KNN classifier is faster for smaller dataset whereas, ANN is a good choice for a scaled dataset. Fig. 6. Shows the graphical user interface (GUI) of our plant identification system.

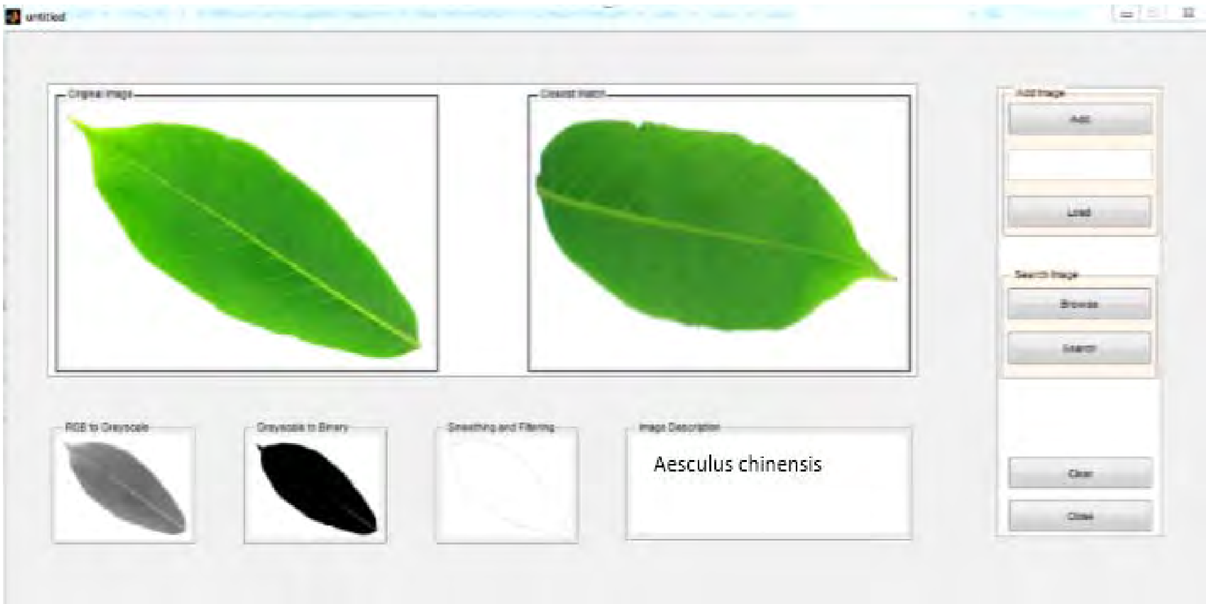


Fig. 6. Snapshot of a sample run of program.

5. Conclusion

In this paper a new robust & computationally efficient system is presented that takes into consideration the color features and tooth features of the leaf in addition to the shape features. We finally used a combination of color, shape, morphological and tooth features. The system was tested on Flavia dataset by using two classifier and the results were admissible as can be seen in experimental results.

The proposed work can be further extended to identify complex images with petiole and clustered leafs and real time images of leaf.

Acknowledgements

We would like to thank the department of Computer Science and Engineering, Amity University, Noida for giving us the resources and the freedom to pursue this project.

References

[1] Arora A., Gupta A., Bagmar N., Mishra S., Bhattacharya A.: A Plant Identification System using Shape and Morphological Features on Segmented Leaflets: Team IITK, CLEF 2012 In: CLEF (Online Working Notes/ Labs/Workshop). (2012)

[2] Neeraj Kumar, Peter N. Belhumeur, Arijit Biswas, David W. Jacobs, W. John Kress, Ida C. Lopez, and João V. B. Soares, "Leafsnap: A Computer Vision System for Automatic Plant Species Identification", Computer Vision – ECCV 2012, 2012, pp 502-516.

[3] Pavan Kumar Mishra, Sanjay Kumar Maurya, Ravindra Kumar Singh, Arun Kumar Misra, "A semi-automatic plant identification based on digital leaf and flower Images", International Conference On Advances In Engineering, Science And Management (ICAESM), 2012, pp. 68-73.

[4] ArunPriya C., Balasaravanan T., 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, 2012, pp. 428-432.

[5] N.Valliammal, Dr. S.N.Geethalakshmi, "Automatic Recognition System Using Preferential Image Segmentation for Leaf and Flower Images", Computer Science & Engineering: An International Journal (CSEIJ), 2011, pp. 13-25.

[6] Dr. H.B.Kekre, Mr. Dhirendra Mishra, Ms. Stuti Narula, Ms. Vidhi Shah, "Color Feature Extraction for CBIR", International Journal of Engineering Science and Technology (IJEST), 2011, pp. 8357-8365.

[7] Javed Hossain, M. Ashraful Amin, "Leaf Shape Identification Based Plant Biometrics", Proceedings of 13th International Conference on Computer and Information Technology (ICCIT), 2010, pp. 458-463.

[8] Jiazhi Pan and Yong He "Recognition of plants by leaves digital image and neural network", International Conference on Computer Science and Software Engineering, 2008, pp 906-910.

[9] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, Qiao-Liang Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", International Symposium on Signal Processing and Information Technology, IEEE, 2007, pp 11-16.

[10] Ji-Xiang Du, Xiao-Feng Wang and Guo-Jun Zhang, "Leaf shape based plant species recognition", Applied Mathematics and Computation 185 , 2007 pp 883-893.

[11] Qingfeng Wu, Changle Zhou, and Chaonan Wang, "Feature Extraction and Automatic Recognition of Plant Leaf Using Artificial Neural Network", Avances en Ciencias de la Computación, 2006, pp 5-12.

View publication stats