Dissertation on
Recognition of leaves using SVM Classifier-An Approach

Thesis submitted towards partial fulfillment
of the requirements for the degree of

Master in Multimedia Development

Submitted by
Debasish Ganguly
EXAMINATION ROLL NO.:M4MMD19001
UNIVERSITY REGISTRATION NO.: 141023 of 2017-2018

Under the guidance of
JOYDEEP MUKHERJEEE

School of Education Technology
Course affiliated to
Faculty of Engineering and Technology
Jadavpur University
Kolkata-700032
India

2019

Master in Multimedia Development
Course affiliated to
Faculty of Engineering and Technology
Jadavpur University
Kolkata, India

CERTIFICATE OF RECOMMENDATION

This is to certify that the thesis entitled **“Recognition of leaves using SVM Classifier-An Approach"** is a bona fide work carried out by **Debasish Ganguly** under our supervision and guidance for partial fulfillment of the requirements for the degree of Master in Multimedia Development in School of Education Technology, during the academic session 2017-2019.

Prof. Matangini Chattopadhyay
SUPERVISOR
Jadavpur University,
Kolkata-700 032

DIRECTOR
School of Education Technology
Jadavpur University,
Kolkata-700 032

DEAN –FISLM

Jadavpur university,Kolkata-700 032

CERTIFICATE OF APPROVAL **

This foregoing thesis is hereby approved as a credible study of an engineering subject carried out and presented in a manner satisfactorily to warranty its acceptance as a prerequisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not endorse or approve any statement made or opinion expressed or conclusion drawn therein but approve the thesis only for purpose for which it has been submitted.

Committee of final examination for evaluation of Thesis

**Only in case the thesis is approved.

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that this thesis contains literature survey and original research work by the undersigned candidate, as part of his **Master in Multimedia Development** studies during academic session 2017-2019.

All information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by this rule and conduct, I have fully cited and referred all material and results that are not original to this work.

NAME:DEBASISH GANGULY

EXAMINATION ROLL NUMBER: M4MMD19001

THESIS TITLE: Recognition of leaves using SVM classifier-An Approach

SIGNATURE: DATE:

ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to my supervisor Joydeep Mukherjee for

his continuous support in my thesis work and related research, for his patience, motivation and

suggestions. Her guidance and valuable suggestions always helped me at time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Masters Degree study. I am very much thankful to her for the motivation and support given during the entire research work.

Besides this, I would also like to thank Prof(Dr.)Matangini Chattopadhyay, Dr. Ranjan Parekh and Dr. Saswati Mukherjee for their continuous encouragement and support during my entire period of study in School of Education Technology. I would like to express my gratitude to the entire staff, lab assistant and research workers specially Mr. Susovan Jana & Mr. Hindol Bhattacharya and all

of those who were associated with my work for their co-operation and opinion. I am very grateful to all my fellow classmates of M. Tech IT (Courseware Engineering) and also Master in Multimedia

Development courses for their helpful suggestions and continuous support.

I would like to thank my parents for always supporting me in every ups and down during my entire period of work. I am also thankful to my friends and well-wishers who always have faith in me.

Date:

Place:

Debasish Ganguly
Class Roll No.:001730401014
Examination RollNo.:M4MMD19001
Master in Multimedia Development
School of education technology,
Kolkata – 700032

TABLE OF CONTENTS

LIST OF FIGURES

page

EXECUTIVE SUMMARY

The current work proposes an approach for the recognition of plants from their digital leaf images using multiple visual features to handle heterogeneous plant types. Recognizing

the fact that plant leaves can have a variety of recognizable features like color (green and non-green) and shape (simple and compound) and texture (vein structure patterns), a single set of features may not be efficient enough for complete recognition of heterogeneous plant types. Accordingly a layered architecture is proposed where each layer handles a specific type of visual characteristics using its own set of features to create a customized data model. Features from various layers are subsequently fed to an array of custom classifiers for a more robust recognition. In this work we enumerate on the color and shape layers only. A dataset involving 450 leaf images divided over 20 classes and including green, non-green, simple and compound leaves, is used to test the performance and effectiveness of the approach.

Chapter 1

1.1 INTRODUCTION

Plants have many uses in industry, medicine, and foodstuff production. Recognizing plant species is an important process to obtain necessary raw materials from correct plants. Plant recognition is also important in environmental protection to correctly observe changes in plant species and population. However, recognizing plants is a difficult task and is generally done by human expert biologists. Designing automatic recognition systems for plants is useful, since it can facilitate fast classification of plants, and have applications in many scientific and industrial fields [10]. For instance, discovery of new species, plant resource surveys, population studies, and plant database management are demanding applications in biology, foodstuff, medicine, and agriculture. Automatic plant recognition may increase efficiency and speed in these fields, save time of human experts, and decrease cost of production stages. 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. Since leaves are considered as important features to characterize plant species, many studies on leaf image retrieval based on shape, venation, color, and texture information have been conducted in computer-aided plant identification systems.

The aim of this work is to develop an approach to classify plants according to leaf features. The classification based on leaf images has an advantage such that sampling leaves (getting photos) is low-cost and convenient. Performance of a leaf recognition system depends on good feature selection and efficient recognition algorithm. In this work, we extended the method in [22] and studied new features and classification algorithms. In addition to shape features, we used color features of leaf images. A collection of machine learning techniques, k Nearest Neighbor, Support Vector Machines, Naive Bayes, and Random Forest, are used to classify both shape and color features. Our results have shown that color features can improve the recognition performance. With Random Forest method, we obtained 96.32% classification accuracy rate in 32 plant species. To the best of our knowledge, the results are state of the art for such a large number of plant types.

1.2 Overview

To identify an item is to recognize the item and associate it with its appropriate name. Such as, the automobile in front of any house is a Honda Accord. Or, a large woody plant in the park is a tree, more specifically a Doug-fir. Identifying a landscape or garden plant requires recognizing the plant by one or more characteristics, such as size, form, leaf shape, flower color, odor, etc., and linking that recognition with a name, either a common or so-called scientific name. Accurate identification of a cultivated plant can be very helpful in knowing how it grows (e.g., size shape, texture, etc.) as well as how to care and protect it from pests and diseases.

First let's look at some **common characteristics of plants that are useful in identifying them**. Now if the same was in a **botany class dealing with plant systematic**, the field of study concerned with identification, naming, classification, and evolution of plants, we would spend a good deal of time on the **reproductive parts of plants**, i.e., mostly the various parts of the flowers, i.e., ovary, stigma, etc. Structural similarity of reproductive parts is an important means by which plants are categorized, grouped, named, and hence identified. However, with many horticultural plants, especially **woody plants, one may have to make an identity without regard to flowers**, for often flowers are not present or are very small, and other characteristics may be more obvious. Some plants characteristics are so obvious or unique that we can recognize them without a detailed examination of the plant.

Pattern recognition is a very important field within computer vision, and the aim of pattern recognition/classification is to classify or recognize the patterns based on extracted features from them. The pattern recognition involves three steps (1) Pre-processing (2) Feature Extraction (3) Classification. In Pre-processing one usually process the image data so it should be in suitable form e.g. one gets an isolated objects after this step. In second step measure the properties of object of interest and in third step, determine the class of object based on features. A brief explanation on the pattern recognition is given in the Figure 4.1.

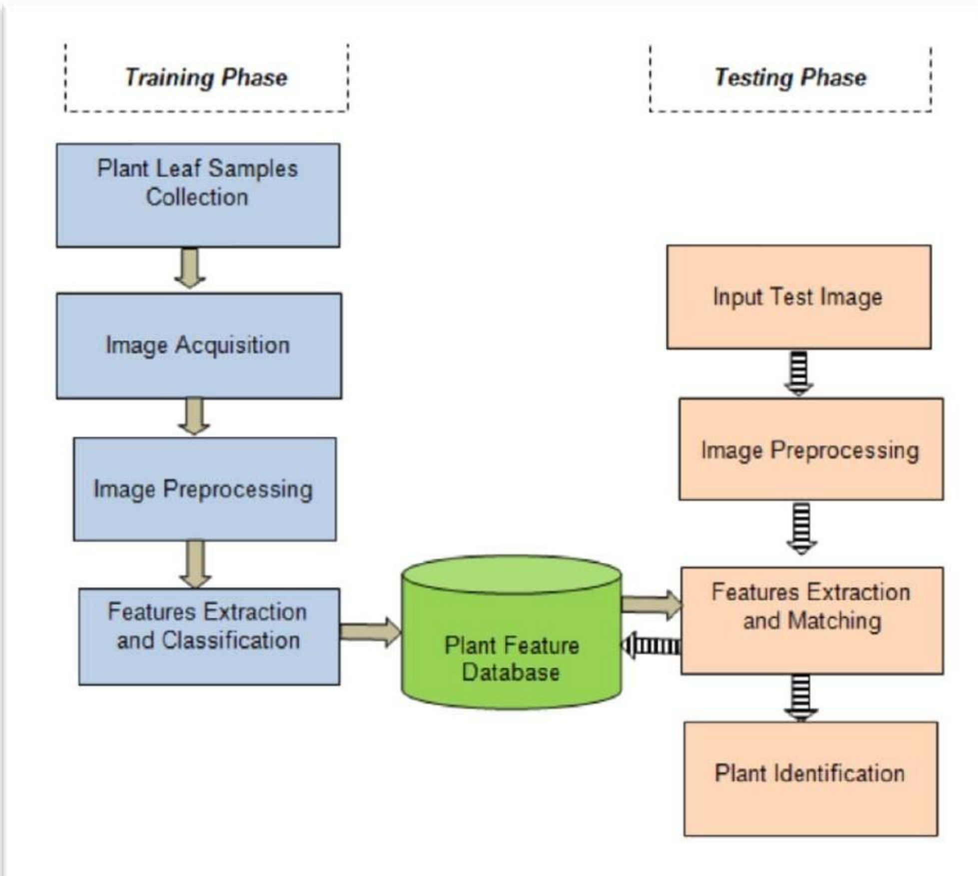


Figure 4.1. Main pattern recognition steps.

IMAGE PREPROCESSING

Before the operations, some of the leaf images are rotated manually for helping the program to arrange leaf apex direction to the right side. Afterwards, automatic preprocessing techniques are applied to all of the leaf images. These preprocessing steps are illustrated on an image as seen in Figure 4.2, while ignoring the color information. As a result, only Gray component for each pixel is computed from the color image by

$$Gray = 0.299 * R + 0.578 * G + 0.114 * B \quad (4.1)$$

Where R, G and B correspond to the color of the pixel [7, 8], respectively.

The rectangle of interest (ROI) of the leaf image should include all the pixels their gray values are smaller than a specific threshold [9], and then the binary image of the leaf is retrieved.

Then the contour of leaf can be extracted.

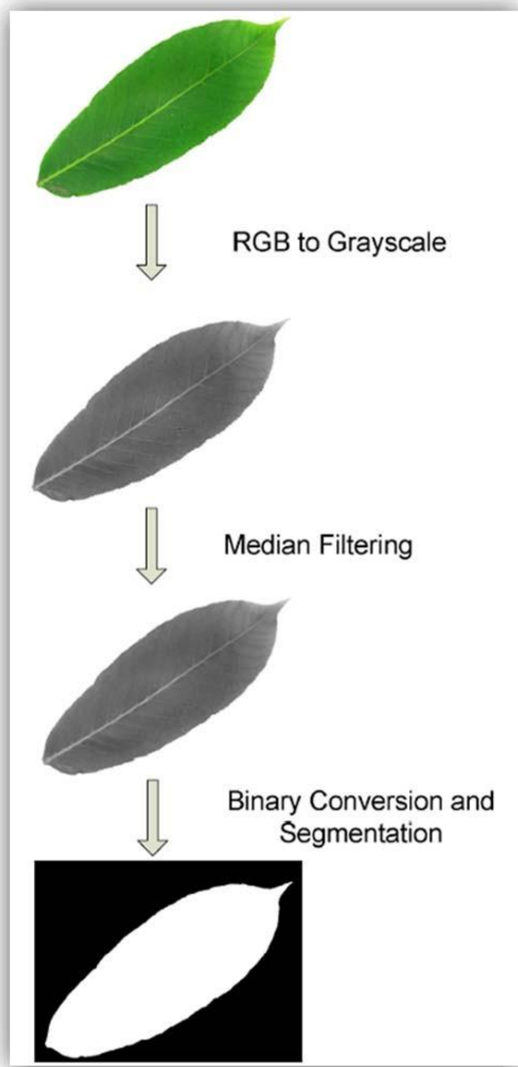


Figure 4.2. Image preprocessing steps.

FEATUREEXTRACTION

After pre-processing, in pattern recognition, the important and essential task is to measure the properties of an object because objects have to be detected based on these computed properties. In the feature extraction step, the task is to describe the regions based on chosen representation, e.g. a region may be represented by its boundary and its boundary is described by its properties (features) such as color, texture, etc.

There are two types of representations, an external representation and internal representation. An external representation is chosen when the primary focus is on shape characteristics. An internal representation is selected when the primary focus is on regional properties such as color and texture. Sometimes the data is used directly to obtain the descriptors such as in determining the texture of a region, the aim of description is to quantify a representation of an object. This implies, one can compute results based on their properties such length, width, area and so on.

- *Area:* Area represents number of pixels in the leaf region. Binary form of our leaf image has black background and white leaf. In this image, number of white pixels represents the area of the leaf.
- *Major Axis:* Major axis is denoted as a line, which lies between apex and base of the leaf.
- *Minor Axis:* Minor axis of the ellipse that has the same normalized second central moments as the leaf region.
- *Perimeter:* Perimeter is the distance around the boundary of leaf region.
- *Convex Hull:* Convex hull represents the smallest convex polygon that encapsulates the leaf region.
- *Minor Axis Length Ratio of Major Axis Length:* This feature is denoted as ratio of minor axis length to major axis length. It is reverse of the aspect ratio that is used in the literature.

CLASSIFICATION(RECOGNITION)

Once the features have been extracted, then these features are to be used to classify and identify an object using SVM classifier to classify plants based on shape-related features of leaf such as aspect ratio, rectangularity, area ratio of convex hull, perimeter ratio of convex hull, sphericity, circularity, eccentricity, form factor and invariant moments.

In general pattern recognition systems, there are two steps in building a classifier: training and testing (or recognition). These steps can be further broken down into sub-steps.

Training:

1. Pre-processing: Process the data so it is in a suitable form.
2. Feature extraction: Reduce the amount of data by extracting relevant information, usually results in a vector of scalar values.
3. Model Estimation: From the finite set of feature vectors, need to estimate a model (usually statistical) for each class of the training data.

Recognition:

1. Pre-processing:
2. Feature extraction: (both steps are same as above)

3. Classification: Compare feature vectors to the various models and find the closest match. One can match the feature vectors obtained in training set.

The algorithm has three main parts: Training, Classification, Segmentation and distance measurement.

1.3 Problem Statement

There are various types of leaves in world, all those leaves cannot be detected manually therefore an efficient procedure using SVM classification has been used to detect those variety of leaves. To overcome the problems of leaves characteristics like shape, texture, etc this SVM classifier is used. To improve the accuracy this method has been used.

- The subtle differences between different species in the same class. Sometime such fine differences can be even challenging to human experts.

- Typically this requires large training data but it is not feasible due to the number of species.

1.4 Objective

Objectives:

1. To introduce plant nomenclature and classification.

2. To become familiar with basic plant morphology.

3. To begin to identify plants using morphological characteristics.

Chapter 2

LITERATURE SURVEY

Wu et al. [1], extracted 12 commonly used digital morphological features which were orthogonalized into 5 principal variables using PCA. They used 900 leaves to classify 30 kinds of plants using a probabilistic neural network system.

Wang et al. [2], employed centroid contour distance (CCD) curve, eccentricity and angle code histogram (ACH).

Fu et al. [3] also used centroid-contour distance curve to represent leaf shapes in which an integrated approach for an ontology-based leaf classification system is proposed. For the leaf contour classification, a scaled Recognition of plants by Leaf Image using Moment Invariant and Texture Analysis CCD code system is proposed to categorize the basic shape and margin type of a leaf by using the similar taxonomy principle adopted by the botanists.

Caglayan et. al [4] Recognizing plants is a vital problem especially for biologists, chemists, and environmentalists. Plant recognition can be performed by human experts manually but it is a time consuming and low-efficiency process. Automation of plant recognition is an important process for the fields working with plants. This paper presents an approach for plant recognition using leaf images. Shape and color features extracted from leaf images are used with k-Nearest Neighbor, Support Vector Machines, Naive Bayes, and Random Forest classification algorithms to recognize plant types. The presented approach is tested on 1897 leaf images and 32 kinds of leaves. The results demonstrated that success rate of plant recognition can be improved up to 96% with Random Forest method when both shape and color features are used.

George et. al.[5] proposed an approach of leaf recognition. They employ multi-layer feed forward network algorithm for leaf recognition using its shape, color and vein features. The leaf images are pre-processed and segmented. The shape features like area, convex area, diameter, length, width, perimeter, eccentricity, solidity, major axis length and Minor axis length are extracted by taking the features from the segmented leaf image. The mean, standard deviation, skewness and kurtosis for red, green and blue color features are extracted. Wiener filtering and canny edge detection are used to identify the vein features v1, v2, v3, v4 at different thresholds by calculating gray level value from the gray level histogram. The features are then exposed to Principal Component Analysis (PCA) for feature reduction and the resultant reduced feature set contains five shape features, six color features and three vein features. The features are then taken as input vectors for multi-layer neural network. Total 150 leaves from Columbia leaf image database are taken as samples. 150 leaves represent 10 different leaf species. The system was trained with 104 leaf images and was tested and validated using 23 leaf images each. The success rate of reduced feature set of leaf shapes and color was 84.7%.

Khmaget et. al[6] proposed The recognition of plants is directly associated to society's life. Leaves from plants are proved to be a feasible source of information used to identify plant species. The recognition system of leaves is accomplished automatically using the experts of human being. Unfortunately, it has their loopholes that are a time consuming processes and low-effectiveness progression. The leaves classification using predictable process is quite complicated, time complexity, and as a result of using very long-termed in botanical science for non-experts that make it more irritated operation. Thus, the prompt developments in digital images, computer vision and object detection and recognition systems encourage scientists to work towards plant species recognition according to image processing technology. In this study, an image processing algorithm in order to find out the shape structure of tested plants is presented. This technique exploits the variant to scaling shift, spin technique, scaling approach, and filtering processes. The leaf contours of the same plants are computed using Support Victor Machine (SVM) where the similar sequences of the same contours usually carry the same features while the different plants sequences have different contours. In this regard, SVM classifier is exploited to be applied as a classifier to the plant's leaf. In the Experiment part, the finding was taken from Flavia dataset and it demonstrated that the suggested technique has high recognition efficiency compared to state of the art methods and is shows better quality images especially in complicated features of digital images such as ridges, edges, lines ,curves and complicated contours.

Sabuet et. al[7] proposed the leaf recognition of 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.

Solano et. al[8] proposed that the number of known and unknown plant species increases as time goes by. Research on plant species can be further advanced if there is a quick and accurate system that can identify plants and hasten the classification process. This system will not only help in accelerating plant classification, but will also allow people who are not morphological experts to conduct their own studies. Leaves is an application designed to classify different plant species based on the leaf's shape and venation. This system uses different image processing and machine learning techniques including centroid-radii, moment invariance, canny edge detection, morphological operations, image difference and artificial neural networks.

Kang et. al[9] proposed that their literature survey focusing on the user interaction aspect reveals that two schemes of image acquisition have been used, one with strong constraint and the other with no constraint. The strong constraint interaction asks users to capture images by placing a leaf on a uniform background such as white paper while the unconstrained interaction allows any form of image capturing. The former one gets a high performance sacrificing the user convenience while the latter one provides a great convenience sacrificing the recognition performance. Our scheme is weakly constrained in the middle of two extremes. The proposed interaction scheme only asks users to center the leaf on smart phone camera screen. The leaf may be on the tree or off the tree. When the leaf is picked off the tree, it is recommended to place it against rather uniform background such as sky, soil, or tree bark.

Sahay et.al [10] proposed that plants are essential resources for nature and people's lives. Plant recognition provides valuable information for plant research and development, and has great impact on environmental protection and exploration. This paper presents a leaf analysis system for plant identification, which consists of three main components. First, given a leaf image, a preprocessing step is conducted for noise reduction. Second, the feature extraction component indentures representative features and computes scale invariant feature descriptors. Third, the matching plant species are indentured and returned using a weighted K nearest neighbor search algorithm. The system is implemented as a Windows phone app and is tested on the Leaf snap dataset, an electronic field guide developed by Columbia University and University of Maryland with different combinations of species at various orientations, scales and levels of brightness.

Lukic et.al[11] proposed that —Leaf recognition is convenient for plant classification and it is an important subfield of pattern recognition. Different leaf features such as color, shape and texture are used as well as different classifiers including artificial neural networks, k nearest neighbor and support vector machines. In this paper we propose an algorithm based on tuned support vector machine as a classifier and Hu moments and uniform local binary pattern histogram parameters as features. Our proposed algorithm was tested on leaf images from standard benchmark database and compared with other approaches from literature where it proved to be more successful.

Chapter 3

Background Study

Despite the importance of the subject of identifying plant diseases using digital image processing, and although this has been studied for at least 30 years, the advances achieved seem to be a little timid. Some facts lead to this conclusion:

Methods are too specific. The ideal method would be able to identify any kind of plant. Evidently, this is unfeasible given the current technological level. However, many of the methods that are being proposed not only are able to deal with only one species of plant, but those plants need to be at a certain growth stage in order to the algorithm to be effective. That is acceptable if the plant is in that specific stage, but it is very limiting otherwise. Many of the researchers do not state this kind of information explicitly, but if their training and test sets include only images of a certain growth stage, which is often the case, the validity of the results cannot be extended to other stages.

The aim of the project is to develop a Leaf recognition program based on specific characteristics extracted from photography. Hence this presents an approach where the plant is identified based on its leaf features such as area, histogram equalization and edge detection and classification. The main purpose of this program is to use MATLAB resources.

Indeed, there are several advantages of combining MATLAB with the leaf recognition program. The result proves this method to be a simple and an efficient attempt.

Chapter 4

Proposed work

Motivation:

However since the world is becoming faster every day, cutting down on time and

increasing the reliability of verification systems, have become an absolute necessity.

People are therefore relying more on automated systems and machine intelligence to perform verification tasks. Image processing and pattern recognition techniques have been employed to perform verification tasks in an automated way by building data models of the signatures and using classifiers to segregate them into pre-defined classes. Most of the existing works on signature recognition focus on modeling the leaf shapes using a variety of features, but these techniques usually do not work satisfactorily when the signatures are transformed using translation, rotation, and scaling.

System Components:

To accomplish the task of recognizing images, an image recognition system must have

the following components:

i) Model database.

ii) Feature detector.

iii)SVM classifier.

The model database consists of all the models that are known to the system. The approach used for the recognition purpose determines the information in the model database. The information can vary from a qualitative or functional description to precise geometric surface information. The object models are abstract feature vectors in many cases. A feature is an object attribute which is important in describing and recognizing an object with respect to other objects. Some of the most commonly used features are color, shape, and size.

The feature detector helps in forming object hypotheses by applying operators to images and identifying locations of features. The types of objects to be identified and the organization of the model database influences the choice of features to be used by a system. Appropriate tools and techniques must be selected by an object classification system for accomplishment of the aforementioned steps. The following are the key issues that should be taken into account while designing an object classification system:

a) Object or model representation:

The objects should be represented in a manner that encapsulates all important information regarding features or attributes without any redundancies in the model database. All the relevant information should be organized in a way that provides easy access by various components of the object classification system.

b) Feature extraction:

Various feature detection issues such as what features should be detected and how the detection can be done reliably, should be considered for feature extraction. Most of the features can be evaluated in two-dimensional images whereas they are related to three dimensional characteristics of objects.

System Design

Acquired images are RGB signature images with 3 color channels. The input leaf image is converted to a binary image in the pre-processing stage and this binary image is pre-processed in different way. Statistical features are then extracted from the images and feature vectors are created. These feature vectors represent the images in the database. Statistical features are kept in two separate databases for training and testing purposes. A classifier is trained using the training features and labels indicating the class of the training Feature vectors. Once trained, the classifier outputs a model which is used to predict the classes for the test images.

Chapter 5: Experiments and Results

Results and Comparison

RESULTS:

Common name	Label	Species samples	accuracy
Acer_pensylvancium	1	35	96
Acer_platanoides	2	20	95
Aesculus_hippocastamon	3	20	94
Amelanchier_canadensis	4	35	95
Asimina_triloba	5	49	96
Ficus_carica	6	45	97
Cornus_florida	7	25	94
Corylus_colurna	8	35	88
Crataegus_phaenopyrum	9	22	93
Ulmus_glabra	10	56	92
Ulmus_parvifolia	11	45	87

Comparison:

Method	Accuracy
Wu(wu et al., 2007)	**93%**
Singh(singh et al., 2010)	**95%**
Kadir(kadir et al.,2011)	**94.7%**
Kulkarni(kulkaeni et al.,2013)	**93.8%**
Proposed method	**96%**

Chapter 6

Support Vector Machine (SVM) Classifier

In machine learning, **Support Vector Machines** (**SVMs**, also **support-vector networks**) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

When data is unlabelled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The **support-vector clustering** algorithm, created by HavaSiegelmann and Vladimir Vapnik, applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabeled data, and is one of the most widely used clustering algorithms in industrial applications.

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyper plane. In other words, given labeled training data (*supervised learning*), the algorithm outputs an optimal hyper plane which categorizes new examples. Let's consider the following simple problem:

We are given a set of n points (vectors) : x1, x2, x3, ….. xn such that xi is a vector of length m, and each belong to one of two classes we label them by "+1" and "-1". So our training setis $(x_1, y_1), (x_2, y_2), \dots (x_n, y_n)$

$$\forall i \quad x_i \in R^m, \ y_i \in \{+1, -1\}$$

Source: https://en.wikipedia.org/wiki/Support-vector_machine

A line is bad if it passes too close to the points because it will be noise sensitive and it will not generalize correctly. Therefore, our goal should be to find the line passing as far as possible from all points.

Then, the operation of the SVM algorithm is based on finding the hyper plane that gives the largest minimum distance to the training examples. Twice, this distance receives the important name of **margin** within SVM's theory. The optimal hyper plane can be represented in an infinite number of different ways by scaling of β and β_0.

Chapter 7

Matching Databases

a. AVAILABLEDATABASES

Caltech University has been working on leaf recognition and has therefore created a dataset consisting of 102 leaf categories and more than 8000 images . The leaves mainly come from USA. Downloads or more information can be found. This database is very attractive since at least 40 images of the same category are present, which is essential for a good recognition at a large scale.

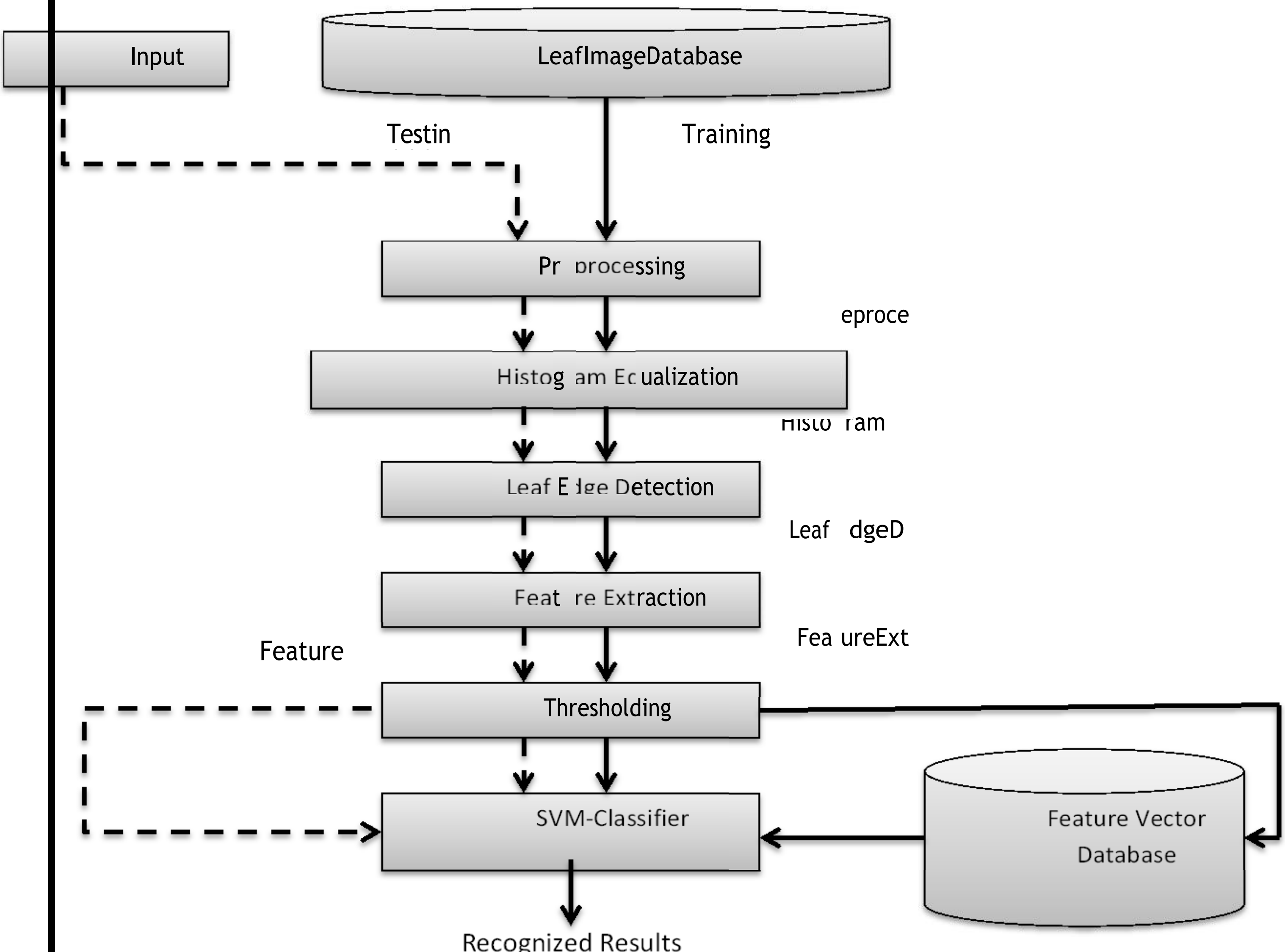
b. RESULTING DATABASE

Finally by mixing a bit the three sources exposed in the previous paragraph, the final database contains 10 images. The repartition is roughly the following 60% from the Caltech dataset, 30% from the other website . The leaves belong to different category.

RESULTS

In order to test the efficiency one can collect additional pictures of flowers present in the database and see if the system recognizes them. But to have significant results another set of suitable test images would have to be found. So a ground truth evaluation of the database has been conducted.

It consists of going through all the images in the database and search the second best match (the first one obviously being the same image). If the leaf image indicated is part of the same category as the leaf under test then it's a successful recognition. By doing this for the whole database, the performance of the system can be evaluated by establishing the recognition rate.



SVM-Classifier

FeatureVector

Database

Chapter 8

Screen shots:

The below list of snapshots Figure 9.1 to Figure 9.13 describe the flow of the project.

Figure 9.1. Snapshot of the menu window.

MENU

The Menu window displays all the image processing steps that has to be carried out on a leaf image.

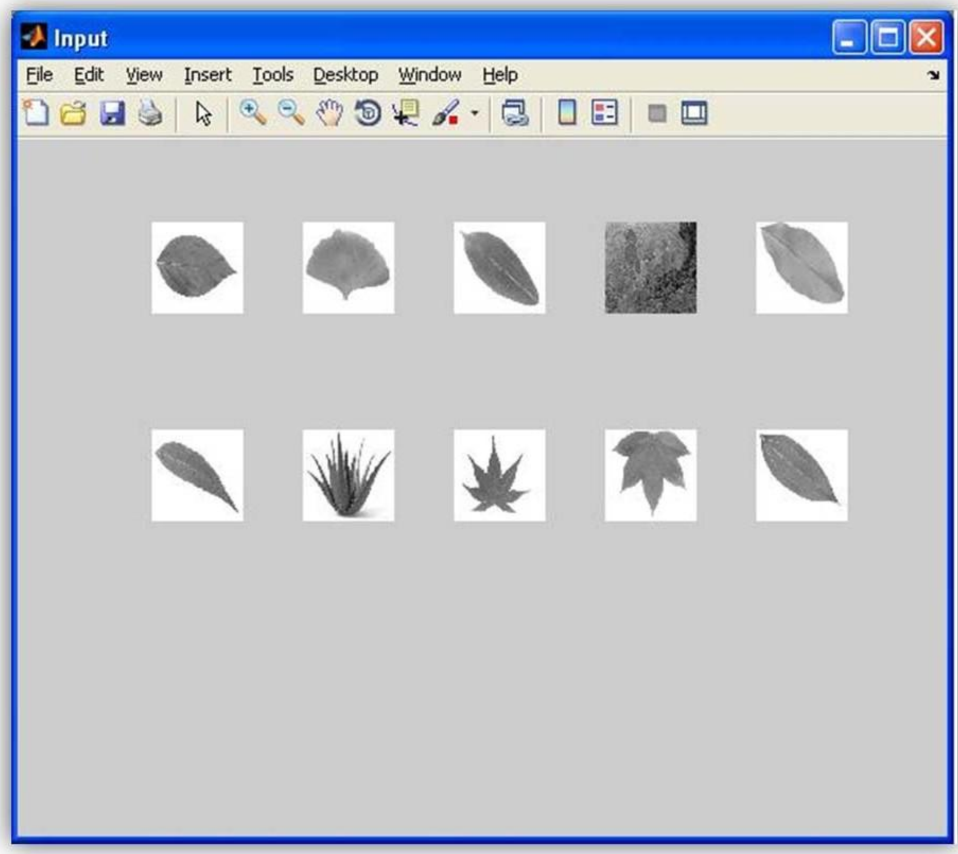


Figure 9.2. Snapshot of the getting images from the database.

Figure 9.3. Snapshot of the leaf detection.

LEAFPROCESS

Leaf process uses the image and performs smoothing on the leaf image.

Figure 9.4. Snapshot of the leaf pre-processing.

Figure 9.5. Snapshot of the histogram equalization.

FEATUREEXTRACTION

The Feature points are extracted from the leaf image.

TESTING

Selection of input leaf image upon which RGB2gray and then smoothing is performed.

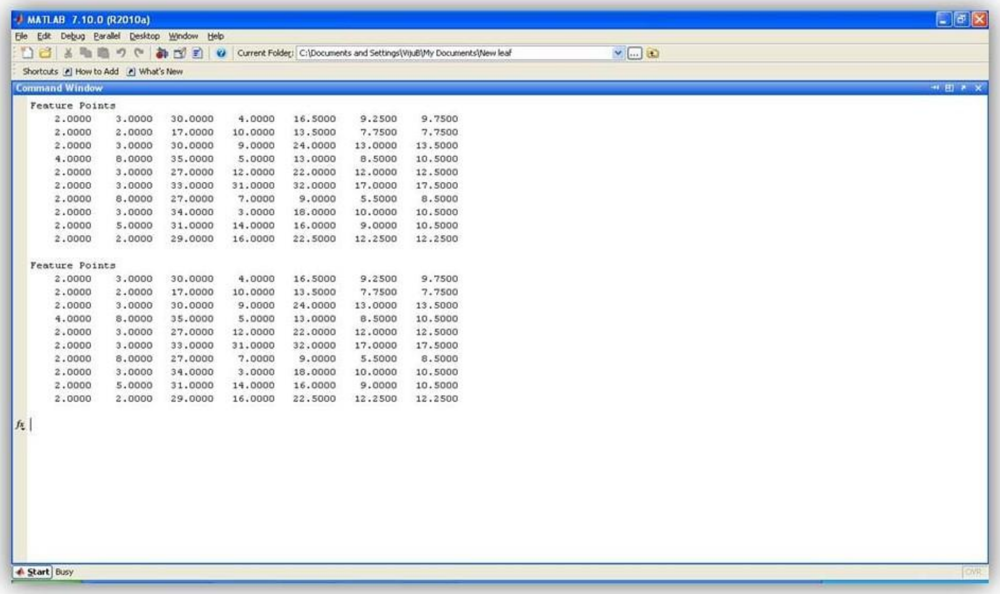


Figure 9.6. Snapshot of feature points retrieved from feature extraction.

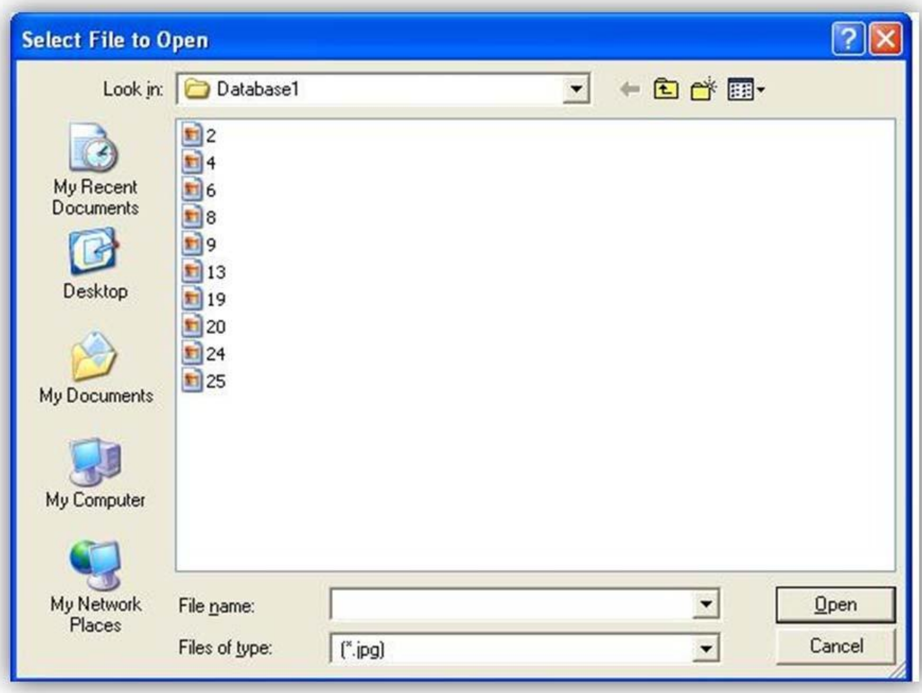


Figure 9.7. Snapshot of the image selection window – To select an image for testing.

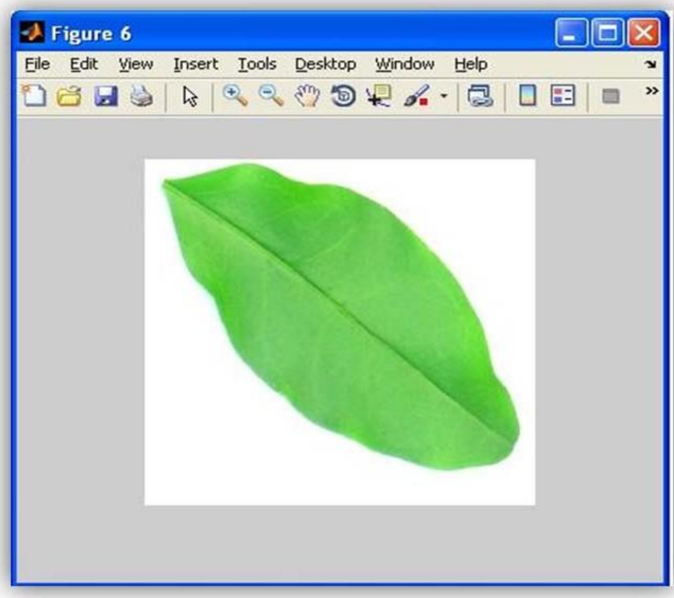


Figure 9.8. Snapshot of the image selected.

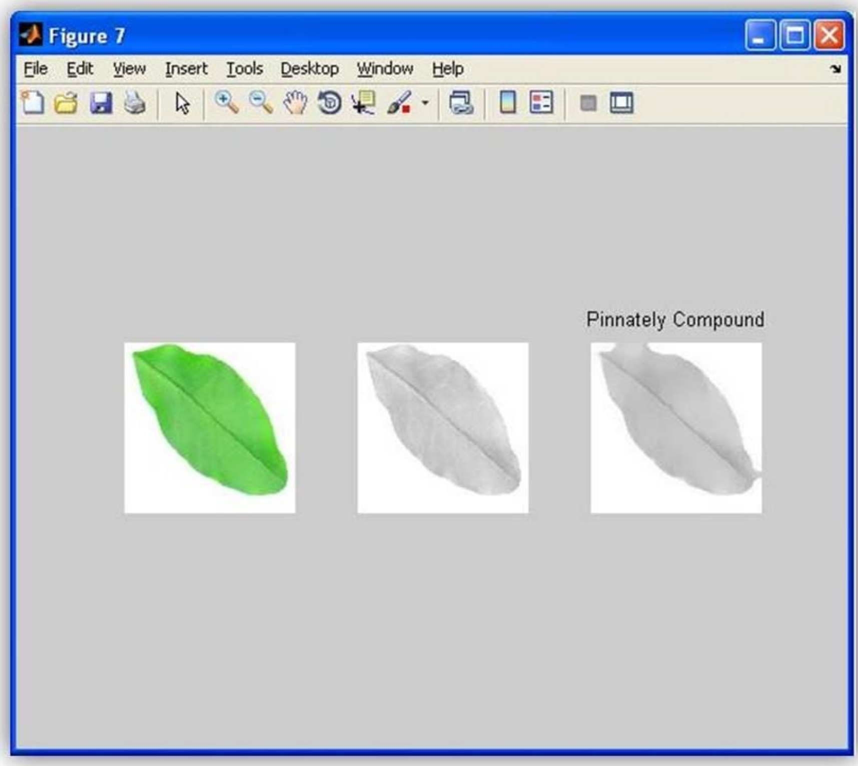


Figure 9.9. Snapshot of the images where the image is converted to RGB2GRAY and then smoothing is performed.

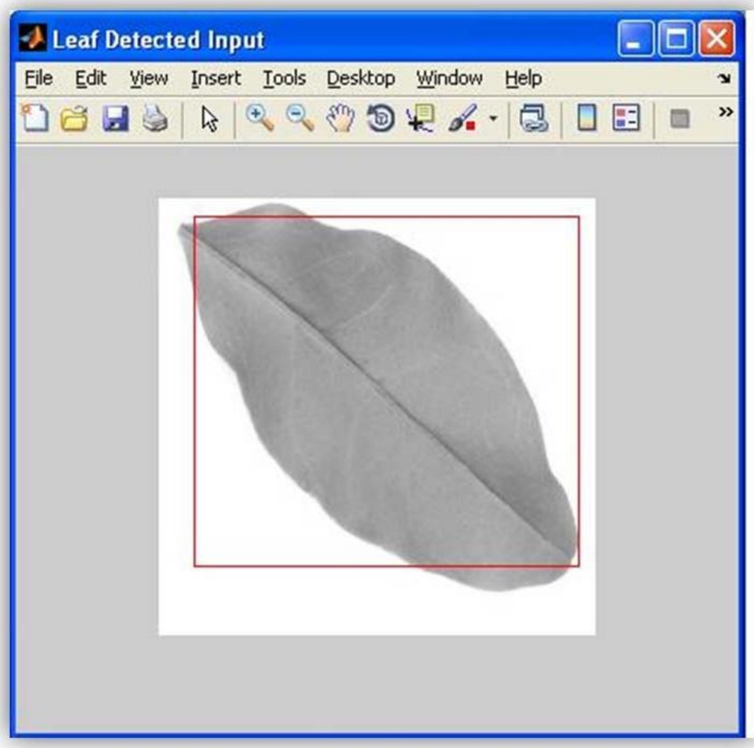


Figure 9.10. Snapshot of the image after applying edge detection.

Figure 9.11. Snapshot of the result window.

LEAF INPUTDETECTED

The input leaf is detected against the leaf image in the database.

Matching displays the result.

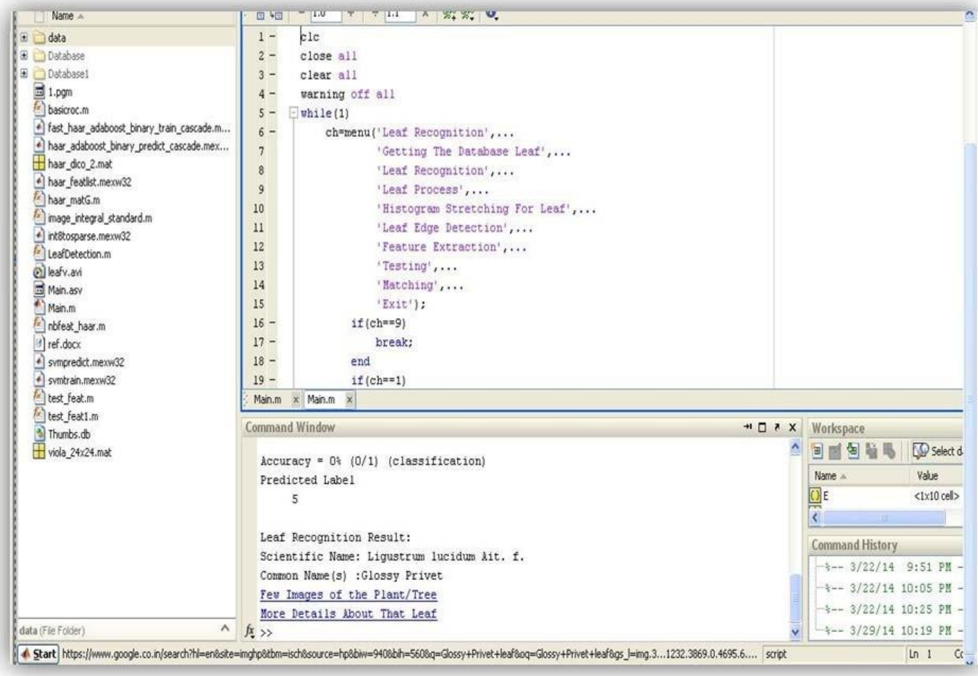


Figure 9.12. Snapshot of the label and Wikipedia link derived from the result.

Chapter 9

Conclusion and Future Work

In this work we have proposed an innovative method the completely automatic identification and classification. of plant species based on leaf recognition, inorder to provide an automated procedure as support for plant cataloging and preserving. Starting from an original leaf image we can process it entirely, in a fully automatic way without need for manual intervention by the user. The proposed method makes use of a new and larger features set, that incorporates shape, color and texture features, extractable from the leaves images easily and quickly. This set of features is used entirely to train a SVM classifier with one rest approach. The experimental results demonstrate that the proposed leaf recognition system has excellent performance, both in terms of accuracy and in terms of speed, less than a second to catalogue an unknown leaf species. The accuracy often reaches 95%and on average is around 93%, so it is better than the methods reported in literature. Further developments could affect the feature extraction phase, in order to tract vein features easy to detect and more discriminating than the features shown in the works reported. in literature. Thereafter, it will be important to test the proposed method on other available datasets that present greater variations in the type of images acquired, both with regard to the dimensions, the resolution and the quality of the images themselves. In this way we would be able to assess the quality of the proposed method and make the changes necessary to maintain a high level of performance and speed. Finally we will be able to focus in the development of a real application for PC, which allows to catalog images of leaves acquired with a digital camera, and mobile application that allows to catalog the plants directly in their natural habitat. This will require the addition of new phases in the algorithm, in particular for the segmentation of the images with the presence of both uniform and non-uniform backgrounds.

Chapter 10

References:

[1] Arribas, J. I., Snchez-Ferrero, G. V., Ruiz-Ruiz, G., and Gmez-Gil, J. (2011). Leaf classification in sunflower crops by computer vision and neural networks.Com-puters and Electronics in Agriculture, 78(1):9 – 18.

[2] Lee, K. B. and Hong, K. S. (2013).An implementa-tion of leaf recognition system using leaf vein and shape. International Journal of Bio-Science and Bio-Technology, 5(2):57–65

[3] Cheng, S.-C., Jhou, J.-J., and Liou, B.-H. (2007). Pdaplantsearch system based on the characteristics of leaves using fuzzy function. In New Trends in Applied Artificial Intelligence, volume 4570 of Lecture Notes in Computer Science, pages 834–844. Springer.

[4] Ehsanirad, A. (2010). Plant classification based on leaf recognition.International Journal of Computer Sci-ence and Information Security, 8(4):78–81.

[5] Haralick, R., Shanmugam, K., and Dinstein, I. (1973).Textural features for image classification. Systems, Man and Cybernetics, IEEE Transactions on, SMC-3(6):610–621

[6] Wu, S., Bao, F., Xu, E., Wang, Y.-X., Chang, Y.-F., and Xiang, Q.-L. (2007). A leaf recognition algorithm forplant classification using probabilistic neural network. In IEEE International Symposium on Signal Process-ing and Information Technology, pages 11–16

[7] Kadir, A., Nugroho, L. E., Susanto, A., and Santosa, P. I.(2011). A comparative experiment of several shape of tobacco leaves. In Congress on Image and Signal Processing (CISP), volume 2, pages 773–776.

[8] Kulkarni, A. H., Rai, D. H. M., Jahagirdar, D. K. A., and Upparamani, P. S. (2013). A leaf recognition technique for plant classification using rbpnn and zernikemoments. International Journal of Advanced Re-search in Computer and Communication Engineering,2(1):82–93.

[9] Pauwels, E. J., de Zeeuw, P. M., and Ranguelova, E. (2009).Computer-assisted tree taxonomy by automated image recognition.Engineering Applications of Artificial Intelligence, 22(1):26–31.

[10] Wu, Q., Zhou, C., and Wang, C. (2006). Feature extraction and automatic recognition of plant leaf using artificial neural network.Avances en Ciencias de la Computacion, pages 5–16.

[11]Zhang, S. and Lei, Y.-K. (2011). Modified locally linear discriminant embedding for plant leaf recognition. Neurocomputing, 74(1415):2284 – 2290.

[12] Juby George, Gladston Raj, "Leaf Recognition using Multi-Layer Perceptron", International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017), 2017.

[13] AsenKhmag, S.A.R Al Haddad, "Recognition System for Leaf Images Based on Its Leaf Contour And Centroid ",Student Conference on Research and Development (SCOReD), 2017

[14] Amala Sabu, Sreekumar K, Rahul R Nair, "Recognition of Ayurvedic Medicinal Plants from Leaves: A Computer Vision Approach ",International Conference on Image Information Processing (ICIIP), 2017

[15] J.Chaki, R.Parekh, S. Bhattacharya, "Plant Leaf Recognition Using a Layered Approach" ,International Journal of Advanced Re-search in Computer and Communication Engineering, 2017.

Chapter 11

Appendix

Segmentation

```
clc;
closeall;
clearall;
% fprint('leaf recognition')

path=' E:\dev\dataset\images\field\output/';
output='gray/';
ext='.jpg';
extn = '.jpg';

NC = 8;    %No. of classes
NT = 10;    %No. of items in training set

%
for CL =  6 : NC
for SA = 1 : NT
class = num2str(CL);
sample = num2str(SA);
    G = imread(strcat(path,class,'(',sample,')',ext));

%    figure;
```

```
%    imshow(I);
% myImage = imread(G);
%  imshow(G);
 I = rgb2gray(G);
%  L = rgb2hsv(I);
%
%
% % %    figure;
% % %    subplot(221), imshow(L);
% %
%    l = L(:, :, 1); % Hue image.
%    a = L(:, :, 2); % Saturation image.
%    b = L(:, :, 3); % Value (intensity) image
% %
% % %    subplot(222), imshow(l);
% % %    subplot(223), imshow(a);
% % %    subplot(224), imshow(b);
% % %
% %
% %
% %
%    t=graythresh(b); % otsu's
% % note that without taking standard deviation
% % we get better results for max. images
%    BW = im2bw(b,t);
% %       figure;
% %   imshow(BW);
%    L=imresize(BW, [500, 500]);
%    R = regionprops(L, 'all');
%
%    BW = ~BW;
% %       figure;
```

```
% %    imshow(BW);
%     BW2 = double(BW);
%
%      CC = bwconncomp(BW2, 8);
%
%     % Compute the area of each component:
%     S = regionprops(CC, 'Area');
%
%     % Remove small objects:
%     L = labelmatrix(CC);
%     BW3 = ismember(L, find([S.Area] >= 10000));
%
%
%     BW2 = double(BW3);
%        for i=(1:3)
%            I(:,:,i) = uint8(double(I(:,:,i)).*BW2);
%        end
% %     figure;
% %     imshow(L);
% % % %
% % % %  Write the output into file
% % % %
% % % % I=rgb2gray(I);
%
imwrite(I,strcat(output,class,'(',sample,')',extn));
%imwrite(I,'output/eg.png');
end
end
```

multisvm

```
function [itrfin] = multisvm( T,C,test )
%Inputs: T=Training Matrix, C=Group, test=Testing matrix
%Outputs: itrfin=Resultant class

itrind=size(test,1);
itrfin=[];
Cb=C;
Tb=T;
fortempind=1:itrind
tst=test(tempind,:);
   C=Cb;
   T=Tb;
   u=unique(C);
   N=length(u);
   c4=[];
   c3=[];
   j=1;
   k=1;
if(N>2)
itr=1;
classes=0;
cond=max(C)-min(C);
while((classes~=1)&&(itr<=length(u))&& size(C,2)>1 &&cond>0)
%This while loop is the multiclass SVM Trick
       c1=(C==u(itr));
newClass=c1;
%svmStruct = svmtrain(T,newClass,'kernel_function','rbf'); % I am using rbf kernel function, you must change it also
svmStruct = svmtrain(T,newClass);
classes = svmclassify(svmStruct,tst);
```

```
% This is the loop for Reduction of Training Set
fori=1:size(newClass,2)
ifnewClass(1,i)==0;
c3(k,:)=T(i,:);
                k=k+1;
end
end
    T=c3;
    c3=[];
    k=1;

% This is the loop for reduction of group
fori=1:size(newClass,2)
ifnewClass(1,i)==0;
c4(1,j)=C(1,i);
                j=j+1;
end
end
    C=c4;
    c4=[];
    j=1;

cond=max(C)-min(C); % Condition for avoiding group
%to contain similar type of values
%and the reduce them to process

% This condition can select the particular value of iteration
% base on classes
if classes~=1
itr=itr+1;
end
end
```

```
end

valt=Cb==u(itr);       % This logic is used to allow classification
val=Cb(valt==1);       % of multiple rows testing matrix
val=unique(val);
itrfin(tempind,:)=val;
end

end

% Give more suggestions for improving the program.
```

Feature extraction

```
function [Feature_Vector] = Extract_FeaturesofLeaves(queryImage)

queryImage = imresize(queryImage, [256 256]);
hsvHist = hsvHistogram(queryImage);
autoCorrelogram = colorAutoCorrelogram(queryImage);
color_moments = colorMoments(queryImage);
% for gabor filters we need gary scale image
img = double(rgb2gray(queryImage))/255;
   [meanAmplitude, msEnergy] = gaborWavelet(img, 4, 6); % 4 = number of scales, 6 = number of orientations
wavelet_moments = waveletTransform(queryImage);
% construct the queryImage feature vector
Feature_Vector = [hsvHistautoCorrelogramcolor_momentsmeanAmplitudemsEnergywavelet_moments];
```

Detect leaves

```
closeall
```

```
clearall
clc

[filename, pathname] = uigetfile({'*.*';'*.bmp';'*.jpg';'*.gif'}, 'Pick an Image File');
Img = imread([pathname,filename]);
Img = imresize(Img,[256,256]);
%figure, imshow(Img); title('Query Image');

% Enhance Contrast
I = imadjust(Img,stretchlim(Img));
figure, imshow(I);title('Contrast Enhanced');

% Extract Features from query image
[Feature_Vector] = Extract_FeaturesofLeaf(I);
whosFeature_Vector

% Load Training Features
load('TrainFeat_Leaf.mat')

test = Feature_Vector;
result = multisvm(TrainFeat,Train_Label,test);
disp(result);

if result == 1
helpdlg(' acer_ginnala ');
disp(' acer_ginnala ');
elseif result == 2
helpdlg(' acer_campestre ');
disp('acer_campestre');
elseif result == 3
helpdlg(' acer_pensylvanicum ');
disp(' acer_pensylvanicum ');
```

```
elseif result == 4
helpdlg(' amelanchier_canadensis ');
disp(' amelanchier_canadensis ');
```