Smart Herbal Plant Reognition System

# System Design Document

Name : M.V.R.Perera

Index: 160477X

Table of Contents

[1. Introduction 1](#_Toc453066740)

[1.1 Purpose of the System 1](#_Toc453066741)

[2. General Overview and Design Guidelines/Approach 2](#_Toc453066742)

[2.1 General Overview 2](#_Toc453066743)

[2.2 Assumptions/Constraints/Risks 2](#_Toc453066744)

[2.2.1 Assumptions 2](#_Toc453066745)

[2.2.2 Constraints 2](#_Toc453066746)

[3. Design Considerations 4](#_Toc453066749)

[3.1 Goals and Guidelines 4](#_Toc453066750)

[3.2 Development Methods & Contingencies 4](#_Toc453066751)

[4. System Architecture and Architecture Design 6](#_Toc453066754)

[5. System Design 10](#_Toc453066768)

[5.1 Business Requirements 10](#_Toc453066769)

[5.2 Database Design 10](#_Toc453066770)

[Appendix A: Acronyms 20](#_Toc453066791)

[Appendix B: Glossary 21](#_Toc453066792)

[Appendix C: Referenced Documents 22](#_Toc453066793)

## Introduction

Smart Herbal Plant Recognition System is a sytem which can recognize the Sri Lankan herbal plants, by taking a picture of a leaf of the selected plant using the mobile phone. A digital plant identification system can be used to quickly characterize the herbal plants without using the help of the botanist who are experts in that area. So this system automize their work and also reduce the time that we spent to find the characteristics of a herbal plant, such as name, uses of the plant, same kind of plants which have the same uses. Although there are many researches done in the area of the leaf identification in inventory systems, most of them are semi-automated systems. Most of the systems I saw are made only using the form or the structure of a leaf. Some have used tooth features also. So it is needed to develop a system which is fully automated and also which needs a least human interaction to identify the plant. So from this document I am expecting to give the details of system design details, that I am trying to develop which is made considering geometric features, structure, form, colour and tooth features in order to increase the efficiency of the system.

### Purpose of the System

There are few main purposes of making this herbal plant recognition system.

* As there is no mobile application to recognize the herbal plants of the Sri Lanka, this will be helpful.
* To make a fully autonomous leaf identification system with least human interaction.
* To increase the efficiency of the leaf recognition, rather than the existing systems.
* To reduce the time and cost that people spend to find a botanist, who can recognize the leaves.
* For studying purposes, who are doing researches in these areas.
* Medicinal purposes for botanists
* For management of herbal plants in agriculture

## General Overview and Design Guidelines/Approach

### General Overview

Leaf of different plants have different characteristics, which can be used to classify them. We have to use digital image processing and, machine vision technology in order to make a simple and efficient leaf classification system. In this method, there are three main phases to complete. They are,

* Pre processing
* Feature extraction
* Classification

**Pre Processing**

It is a technique used to enhance the digital images, before using them for computational processing.

**Feature Extraction**

In this phase, features of the leaf are gained using the color and the shape of the leaf image. These features are used as the inputs for the classification

**Classification**

Above found features are used for efficient classification and tested and compared using the Artificial Neural Network(ANN) and Euclidean Classifier(KNN).

So the main steps of the proposed methodology for this project is, described by the following diagram.

Image Acquisition

Preprocessing

Feature Extraction

Classification, Training and Testing

Final Result

* Image Acquisition

An image of leaf can be easily taken from the mobile phone. Image can be of any size. But the application would give better results, if the leaf image is taken in a single colour background with no stem.

* Preprocessing

Before extracting any specific information, image processing steps are carried out. After that actual image anlaysis will be done. Image have to be changed in an acceptable way such as eliminating the noise and correct the distorted or disgraded data. So we have to use techniques such as, grayscale conversion, binarization, smoothing, filtering and edge detection which are used for enhancement of the leaf image. Following steps will show the preprocessing techniques used for an example leaf.



* Feature extraction

Our method takes into account, the color and shape features of the leaf. Because there may be different plants with same color or same shape, by considering only one aspect, it may give wrong results.

**Color features**

At first, the three color planes of the leaf is separated, which are named as red, blue and green. And then for each plane, and for each color, row mean and column mean is calculated. After that, for each plane, average of row mean and average of column mean is calculated. Features taken from all three planes will be taken to make a feature vector. Those feature vectors made for images, will be stored in a feature database.

**Shape features**

The geometrics features, that I am using to detect the shape of the leaf are, diameter, physiological length, physiological width, leaf area and leaf perimeter.

(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.



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 4πA/P2.

(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

* Classification, training and 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*,

*ai* 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 → 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 *fs* but also a confidence vector describing the probability that a given sample belongs to a certain class.

**Classifier**

To approaches will be followed to classify the dataset. They are the neural network and Euclidean Distance Method. A Neural network have 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

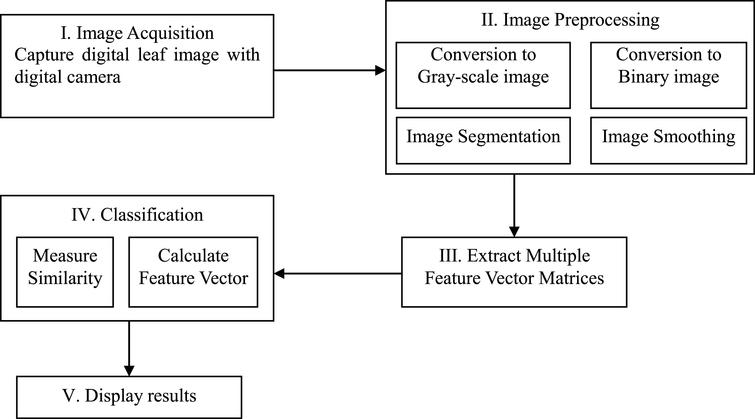
passing from one unit to another, and it is these weightings which are tuned in the training phase to adapt a neural network to the particular problem at hand. This is the learning phase.

Neural Network Process is given below.

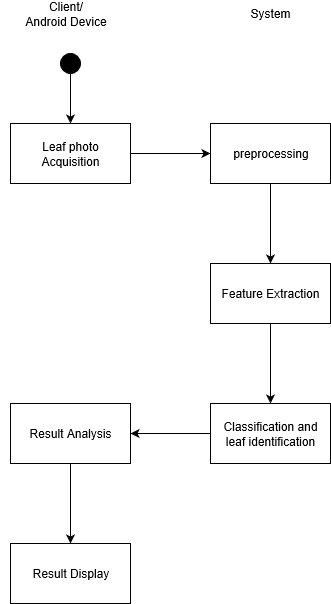


**Training and Testing**

* Block Diagram for the above mentioned Steps



* Activity Diagram



.

### Assumptions/Constraints/Risks

#### Assumptions

User must have an experience of using a mobile phone with its camera and application.

#### Constraints

In this methods I have used leafs to identify the plants, instead of using fruits, flowers, root or stem characteristics. Reason for that is, fruits and flowers may not have with herbal plants and also, those are seasonal. And root and stem characterists of a plant is changed from plant to plant.

User have to take the image of the leaf, in a single color, untextured background and should not be without the stem for best results. The leaf have to be a Sri Lankan Herbal Plant in order to get the correct result.

## Design Considerations

### Goals and Guidelines

Having a considereable accuracy of image classifying is a goal for this project. For that we have to use both color and shape features of the leaf. That also increases the efficiency of the system.

And the accuracy of the classifiers also affects.

**Classifier Accuracy (%)**

* KNN 85.9
* ANN 93.3

Main goal is to make the leaf identification system is efficient rather than the existing systems.

For coding, it can be implemented using the language python.

Main focus will be on the following points.

* Performance

Improving the efficiency of the currently available systems.

* Portability

As this is compatible with android mobile phones, portability is maintained. But computational-wise it should not reduce the performance, because of portability.

* Robustness

Small variations in leaf shape, photographing style and background should not affect the outcome of the system.

plant species classification is based on the nearest neighbor distance of the query leaf’s features from the median features of each species in the training set. Our method proved quite robust under reasonable conditions.

### Development Methods & Contingencies

Yet in the planning stage.

## *C:\Users\Vinura\Desktop\automatedHerbal.PNG*System Architecture and Architecture Design

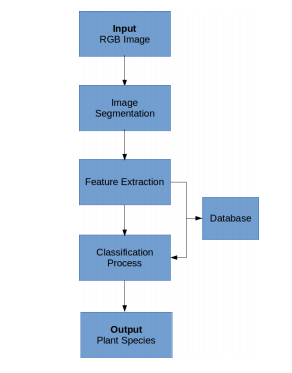
## System Design

### Business Requirements

In this system user has an mobile application in his/her mobile phone, and used it to take a picture and open via the application and send it for processing to identify. We have to consider whether this is an online/offline application considering the requirements and memory required and database access method.

### Database Design

Database will be accessed as seen from the following diagram.



Prior to the classification of a new and unknown specimen, it is necessary to possess a database of features for all the different specimens in each species to train statistical classifiers. This numeric database is derived from an image database.The process ends with the application of a classification process to the features extracted from the specimen to be classified. These are compared to references in the database and an output with the probable species to which the input leaf belongs is given.

Appendix A: Acronyms

KNN

ANN

Appendix B: Glossary

* Acquisition : an asset or object bought or obtained, typically by a library or museum

Appendix C: Referenced Documents