HERB LEAVES RECOGNITION

NUR SYAFIQAH FARHAH BT AZIZ

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS TESIS

JUDUL: HERB LEAVES RECOGNITION

SESI PENGAJIAN: 2017

Saya _____NUR SYAFIQAH FARHAH BT AZIZ_____

(HURUF BESAR)

mengaku membenarkan tesis (PSM/~~Sarjana/Doktor Falsafah~~) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ** Sila tandakan (/)

_____ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

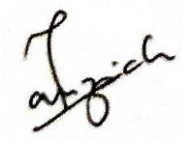
_____ TERHAD (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

_____ TIDAK TERHAD

_____
(TANDATANGAN PENULIS)

Alamat tetap: PT.146, JALAN TO'UBAN,
KAMPUNG KUBANG BATU,
17050 PASIR MAS,
KELANTAN

Tarikh: Tarikh: __30/8/2017_____



_____
(TANDATANGAN PENYELIA)

PUAN FAUZIAH BT KASMIN
Nama Penyelia

Tarikh: Tarikh: _30/8/2017__

CATATAN: *Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM)

** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa

HERB LEAVES RECOGNITION

NUR SYAFIQAH FARHAH BT AZIZ

This report is submitted in partial fulfilment of the requirement for the Bachelor of Computer Science (Artificial Intelligence)

FACULTY OF INFORMATION AND COMMUNICATION AND TECHNOLOGY
UNIVERSITI TECHNIKAL MALAYSIA MELAKA

2017

DECLARATION

I hereby declare that this project report entitled

HERB LEAVES RECOGNITION

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT: _____ Date: 30/8/2017

(NUR SYAFIQAH FARHAH BT AZIZ)

SUPERVISOR: _____ Date: 30/8/2017

(PUAN FAUZIAH BT KASMIN)

DEDICATION

I would like to show my gratitude to my supervisor, Puan Fauziah Bt Kasmin, from Universiti Teknikal Malaysia Melaka (UTeM) for always guiding and helping me to complete up this Final Year Project throughout numerous consultations. Most important, I would like to show my highest gratitude to my beloved family and my lecturers who always support me and continuously keeping up my spirit. Once again sincerely thank you all for have been supportive and encouraging me to complete this thesis and through the year of my studies. I would like to expand my deepest gratitude to all those friends who have directly or indirectly guide me in completing this project.

ACKNOWLEDGEMENT

In performing my project, I would like to take an opportunity to express my gratitude to those who help and show their kindness in assisting me to complete this project. First and foremost, I would like to show my gratitude to my supervisor, Puan Fauziah Bt Kasmin, from Universiti Teknikal Malaysia Melaka (UTeM) for all guidance and encouragement. I am in debt for the patient, support and motivation from the beginning to the end of my project. Thank you for instructing and teaching me in developing the understanding of the project and eventually me enable to grasp the essence of this project.

Most important, I would like to show my highest gratitude to my beloved family who always support me and continuously keeping up my spirit. Once again sincerely thank you all for have been supportive and encouraging me to complete this thesis and through the year of my studies.

I would like to expand my deepest gratitude to all those friends who have directly or indirectly guide me in completing this project.

ABSTRACT

Herbs are any plants used for food, flavoring, medicine, or fragrances for their savory or aromatic properties. Culinary use typically distinguishes herbs from spices. Herbs refers to the leafy green or flowering parts of a plant (either fresh or dried), while spices are produced from other parts of the plant (usually dried), including seeds, bark, roots and fruits. The purpose of this project is to develop the herbs' leaves type recognition system. The known of herbs become less conscious among people nowadays, by using this herb's leaves type recognition it will ease the manpower to recognize the leaf type. Recognition of plants is a challenging computer vision problem that requires dealing with irregular shapes and textures with high intra class variability. The main leaves features considered in this project are leaves major length, minor length, and area. To complete this project, the method decided to use is image processing method which is to detect the herbs' leaves type by using MATLAB R2012b software. This project used K-Nearest Neighbors classifier (KNN) and Linear Discriminant Analysis classifier (LDA) to classify the dataset. The purpose of using the KNN classifier is because this classifier is robust to noisy training data. Thus, it is more effective for large training data. The purposed of LDA method is to improve the discriminative ability of projected samples. The dataset for this project are divided into two which are for testing set and training set, 60 and 180 dataset respectively. The images captured are grouped according to their types. The expected outcomes for this project is the system can recognize the herb leaves type correctly according to their classes and groups.

ABSTRAK

Herba adalah tumbuhan yang digunakan untuk makanan, perisa, perubatan, atau wangian untuk keselesaan atau aromatik mereka. Kegunaan herba untuk masakan biasanya berbeza dari rempah. Herba merujuk kepada bahagian tumbuhan yang hijau atau berbunga (sama ada segar atau kering), manakala rempah dihasilkan dari bahagian tumbuhan yang lain (biasanya kering), termasuk biji benih, kulit kayu, akar dan buah-buahan. Tujuan projek ini adalah untuk menghasilkan sistem pengecaman jenis daun herba. Herba menjadi kurang dikenali di kalangan rakyat pada masa kini, dengan menggunakan pengecaman jenis daun herba ini ia akan mengurangkan tenaga kerja untuk mengenali jenis daun. Pengecaman tumbuhan adalah masalah yang mencabar bagi penglihatan komputer yang memerlukan berurusan dengan bentuk dan tekstur yang tidak teratur dengan kepelbagaian kelas intra yang tinggi. Ciri utama daun yang digunakan dalam projek ini adalah panjang utama daun, panjang kecil daun, dan luas daun. Untuk menyiapkan projek ini, kaedah yang diputuskan untuk digunakan adalah kaedah pengecaman imej untuk mengecam jenis daun herba dengan menggunakan perisian MATLAB R2012b. Projek ini menggunakan pengelasan K-Nearest Neighbor (KNN) dan pengelasan Linear Discriminant Analysis (LDA) untuk mengelaskan set data. Tujuan menggunakan pengelas KNN adalah kerana pengelas ini adalah kukuh untuk data latihan bising. Oleh itu, ia adalah lebih berkesan untuk data latihan besar. Tujuan kaedah LDA adalah untuk meningkatkan keupayaan diskriminasi sampel yang diunjurkan. Set data untuk projek ini dibahagikan kepada dua iaitu masing-masing bagi set ujian dan set latihan, 60 dan 180 set data. Imej yang diambil dikumpulkan mengikut jenis. Hasil yang dijangka untuk projek ini adalah sistem dapat mengecam daun herba dengan betul mengikut kelas dan kumpulan mereka.

TABLE OF CONTENT

LIST OF TABLE

LIST OF FIGURE

CHAPTER I

INTRODUCTION

1.1 Project Background

Recognition of plants is a challenging computer vision problem that requires dealing with irregular shapes and textures with high interclass variability. Interest in methods for visual classification of plants has grown recently as devices equipped with cameras became ubiquitous, making intelligent field guides, education tools and automation in forestry and agriculture practical. Leaf is a part of an importance agent in our daily life. Every green leaf is making process of photosynthesis which mean the diffusion of gas such as carbon dioxide and absorb the sunlight energy to make food. Then it will remove the oxygen as a waste product for plant which is most living creatures need in order to breath. The oxygen is released through the stomata into the air, giving us what we need to continue our life. For this project, it is focused on the type of herbs leaves. The herbs leaves was selected as a project scope. Herbs are any plants used for food, flavoring, medicine, or fragrances for their savory or aromatic properties. Culinary use typically distinguishes herbs from spices. Herbs refers to the leafy green or flowering parts of a plant (either fresh or dried), while spices are produced from other parts of the

plant (usually dried), including seeds, bark, roots and fruits. To complete this project, the method decided to use is image processing method which is to detect the herb leaves type by using MATLAB software. The main features of herb leaves considered in this project are leaves major length, minor length, and area. This project used K-Nearest Neighbor classifier (KNN) and Linear Discriminant Analysis Classifier (LDA) to classify the dataset. The purpose of using the KNN classifier is because this classifier is more effective for large training data. The purposed of LDA method is to improve the discriminative ability of projected samples. The data that will be taken are curry leaf, lime leaf, kaffir lime leaf, turmeric leaf, *pandan* leaf and galangal leaf. The dataset for this project are divided into two which are for testing set and training set, 60 and 180 dataset respectively. The objectives of the project are to collect the data of herb leaves images for the database, to extract the features of the images of herb leaves and to classify the herb leaves according to their classes and groups. The expected outcomes from this project is can create database, successful extract the features of the types of leaves, and can show the correct class of leaves.

1.2 Problem Statements

There are plenty of herb types around us in our daily life. The herbs used to be popular around the old generation used for food, medicine, flavoring, or fragrances for their savory or aromatic properties, even used in some spiritual cases. Days by days the old generation are leaving us, most of the nowadays people does not been exposed to the importance of the herbs in daily live. The herbs become unpopular among the new generation, the problem in recognition of herbs' leaves type become serious. Most of the people does not know what the types of the herbs are, they just do not care about it. Recognition of plants is a challenging computer vision problem that requires dealing with irregular shapes and textures with high interclass variability. The lack of research about

the herb types, makes the people difficult to know more about it. By doing this project, it will help the people to know the leaves types with the easier way. This project will ease peoples to recognize the leaves types, by using this herb leaves recognition system. By using this system it will recognize the leaves types and will classify it according to their group. This technology can help people to known the herbs types easily.

1.3 Objectives

This project embarks on the following objectives:

1. To collect the data of herb leaves images for the database.
2. To extract the features of the images of herb leaves type.
3. To classify the herb leaves according to their classes and groups.

1.4 Scope

In this project, the scopes have been divide into three parts which are data scope, software scope, and the study scope. The data scope in this project is, the data that will be used are manually collected from 6 types of herbs leaves around of us. The leaves images are captured by using camera and the features of the images are extract as data. The data that will be taken are curry leaf, lime leaf, kaffir lime leaf, turmeric leaf, *pandan* leaf and

galangal leaf. The total of the images (data) are 180 which are curry leaves, 30, lime leaves, 23, kaffir lime leaves, 30, turmeric leaves, 33, *pandan* leaves, 34, and galangal leaves, 30 respectively. The data for test dataset, and train dataset are 60 and 180 respectively. The main software scope used to complete this project is MATLAB R2012b. This software is used to extract the leaves features and classify the leaves data. Other than MATLAB R2012b, the Microsoft Word 2013 is used to complete the report for this project. The study scope for this project is, the herbs that grown around us are taken according to the specific types of this project.

1.5 PROJECT SIGNIFICANCE

This project will be benefit to the community and the government. This is due to the technology of image processing are used. This project is to recognize the herb leaves types. This will help the community to recognize the herbs' types easily by using this image processing technique. For the government, this project can be use by the agriculture company such as Ministry of Agriculture and Agro-based Industry abbreviated MOA, Rubber Industry Smallholders Development Authority (RISDA), and other industry to implement this herb leaves type recognition into their agriculture technique. This will ease them to recognize the leaves types and will low the manpower needed to recognize of each leaves type.

1.6 EXPECTED OUTPUT

The expected output from this project is the database for the leaves can be created. The system can extract the features of the leaves, thus it can recognize the herb leaves type correctly according to their class and group. The images captured are grouped according to their types. The extract data from the images then can be classify into the train dataset, 180 dataset and the test dataset, 60 dataset. This will enable the system to train and test the data. The system learn the data by training and testing the dataset, and this will lead to successful identify the types of leaves according to their features. Thus, the system can show the correct class of leaves and the features of the leaves.

1.7 CONCLUSION

For this project, it is focus on the herb leaves recognition. The main features considered in this project are leaves major length, minor length, and area of the leaves. To complete this project, the image processing method will be used to detect the herb leaves type by using MATLAB R2012b software. This project used KNN classifier and LDA classifier to classify the data. The dataset for this project are divided into two which are for testing set and training set, 60 and 180 dataset respectively.

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

In this chapter Literature Review is about finding or studies from the past research or journal which focus on the method used, discussion part and the final result that are similar or related to the project.

According to Trishen Munisami *et al.* (2015) recognition of plants by using a machines requires the use of specialized programs or libraries. Since it is usually need to perform image-preprocessing technique to extract visual information and compare it to the existing set of data.

Based on Trishen Munisami *et al.* (2015) human can undoubtedly recognize distinctive object by using their sense, but machine need the use of sensors and cameras to emulate some sense of humans have.

2.2 Leaf Features and Extraction

Based on Mohamad Faizal Ab Jabal *et al.* (2013) there are 2 categories of features which are geometric and visual feature. The geometric feature usually something that can physically touched and manually measure such as the shape of the leaf, perimeter, length, and diameter. While the visual feature is something that can be measured by using a special method and is not tangible. For instance the visual features are color, vein, and texture of the leaf.

Based on ArunPriya C. *et al.* (2012) proposed the feature extraction uses 12 common of Digital Morphological Features (DMFs), derived from 5 basic features, so that computer can obtain feature values quickly and automatically. The 5 basic features mean are the diameter, physiological length, physiological width, leaf area, and leaf perimeter. From the basic features it will derived into 12 digital morphological which are smooth factor, aspect ratio, form factor, rectangularity, narrow factor, perimeter ratio of diameter, perimeter ratio of physiological length and physiological width, and the vein features.

According to Khatere Meshkini and Hassan Ghassemian (2017), Gray level co-occurrence matrix (GLCM) is widely used texture analysis algorithm because of it is easily implemented. It is also contains the information about the position of pixels having similar gray level values. Based on Khatere Meshkini and Hassan Ghassemian (2017), the GLCM has a great potential for increasing the classification rate of textured images.

2.3 Leaf Classification Accuracy

According to Meeta Kumar *et al.* (2012) the survey on techniques for plant leaf classification have been done. One of the techniques is K-Nearest Neighbors classifier. K-Nearest Neighbor classifier calculates the minimum distance of a given point with other points to determine its class. This is mean it should decide to which class the testing dataset belongs from the training dataset.

Based on Meeta Kumar *et al.* (2012) the k-nearest neighbor's algorithm is amongst the simplest of all machine learning algorithm. Also stated that it is reasonable to assume that objects which are close together (according to some appropriate metric) will belong to the same category.

Besides, according to ArunPriya C. *et al*. (2012), SVM classifier is adopted for the classification approach as it has better accuracy, fast training speed and simple structure. Thus, compared with k-NN method, the proposed algorithm produces better accuracy and takes very less time for execution. Unfortunately, the SVM method is not

suitable for this herb leaves recognition, because this MATLAB2012b software only allow for binary class only while this project have 6 classes. This classifier might can be use if using other software or the other MATLAB version.

2.4 K-Nearest Neighbor (KNN) Classifier

Based on A.Moosavian *et al.* (2013) KNN classifier is a simple classification for non-parametric method. It is perform very well despite of the simplicity of the algorithm. The KNN rule holds the position of training samples and their class, the distance between query data and training sample is being calculated when the decision about new incoming data is needed.

According to Hamid Parvin *et al.* (2010) KNN classification is one of the most fundamental and simple classification methods. The classification is made by examining the label most frequently represented among the K nearest samples on the KNN and taking a vote.

2.5 Linear Discriminant Analysis

The Fisherface method is generally known as Linear Discriminant Analysis (LDA). According to Maryam Imani and Gholam Ali Montazer (2017), the function of LDA is to involve classes' information of pattern by maximizing the ratio of the determinant of between-class scatter matrix and within-class scatter matrix. Moreover, LDA can well separate the classes in a low dimensional subspace, but fails to work in small sample size problems when matrix become singular due to the high dimension of vectors and low number of observations.

Based on Lei Pan *et al.* (2017), the purposed of LDA is to maximize the trace ratio of between-class and within-class scatter matrices. Besides, the LDA can improve the discriminative ability of projected samples. However, the LDA is end to fail when the observed data has multimodel distribution.

2.6 Related Work

In the research paper written by Jeselle Petrize M.Sosa *et al.* (2009) stated that the quickest and easiest way to study the different parts of plant by analyzing its leaves. Plant leaves have numbers of properties namely color, texture, structure or shape. Thus, most algorithms use is the leaf shape because the given input, which are images of the leaf, the shape is the property that can be easily extracted. In this herbs' leaves type recognition project also used the different shapes of leaves, but it is not the main features used to classify the types.

Based on Trishen Munisami *et al.* (2015) the feature extraction used in the project are convex hull information, morphological information, distance maps, and color histogram. The convex hull is using the boundary points of the leaf to extract the number of vertices and the area and perimeter also calculated. The morphological is where the length and the width of leaf is calculated by finding the minimum and maximum x and y coordinates. The distance maps is the finding of vertical and horizontal maps, and the centroid radial map. The color histogram is for cropped part of image, the length and width of bounding box are used as marker to crop the central of the leaf image.

Based on Sofiene Mouine *et al.* (2012) the features matching is done by an approximate similarity search technique based on a Locality Sensitive Hashing (LSH) method. This method is to determine the similarity between two feature vectors by using hash function to reduce the search and the cost time. Compare to this project, the method used to determine the similarity between leaves is by using K-NN classifier.

Another related research was written by Kue-Bum Lee and Kwang-Seok Hong (2013) which is the Fast Fourier Transform (FFT) method is used to classify the leaves vein. The FFT is performed by using the distance feature between contour and centroid on the detected leaf image. The leaf direction can be decide by using leaf vein, and the use the frequency domain data by using FFT.