UNIVERSITI TEKNOLOGI MARA

AGARWOOD CLASSIFICATION BASED ON ODOR PROFILE USING INTELLIGENT SIGNAL PROCESSING TECHNIQUE

MUHAMMAD SHARFI BIN NAJIB

Thesis submitted in fulfillment
of the requirements for the degreee of
Doctor of Philosophy

Faculty of Electrical Engineering

June 2012

ABSTRACT

This thesis presents the classification of Agarwood from Malaysia and Indonesia regions based on signal processing technique. Signal processing for the Agarwood classification is a new area and has yet been actively implemented. In this thesis, the Agarwood has been pre-identified by experts using 32 sensor arrays to measure the Agarwood odor profile. General Agarwood pattern has been plot in 2D diagram. The odor profile from different samples have been normalized and pre-processed and visualized in 3D and 2D plot to find unique patterns. The variation of patterns that has been visualized has been marked as different group samples. From 32 data sensor arrays, several significant data sensor array have been pre-processed using principal component analysis (PCA) as data reduction process. The selected data from PCA are applied as input to compute sensor centroid for k-NN and ANN model design. To test the robustness of the classification techniques, the data sets are randomized for both k-NN classifier and ANN model. The classification results of the k-NN classifier and the ANN model utilizing significant sensor centroid new features for Agarwood grades and regions. It was found that the k-NN classifier and the ANN model is able to classify 100% of Agarwood grade and region.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

Abbreviation

2D	-	2-Dimensional
3D	-	3-Dimensional
AI	-	Artificial Intelligence
ANN	-	Artificial Neural Network
ASP	-	Advance Signal Processing
BPNN	-	Backpropagation Neural Network
CPN	-	Cohonen Propagation Neural Network
C_n	-	No. of Centroid
DM	-	Data Measured
EO	-	Essential Oil
FE	-	Feature Extraction
FFT	-	Fast Fourier Transform
F_n	-	No. of Feature
FRIM	-	Forest Research Institute Malaysia
GC-MS	-	Gas Chromatography/Mass Spectrometry
GC	-	Gas Chromatography
GC-O	-	Gas Chromatography-olfactomaetry
GC-MS-FID	-	Gas Chromatography Flame Ionization Detector
k	-	k-NN variable
k-NN	-	k-nerarest Neighbour
LDA	-	Linear Discriminant Analysis
MS	-	Mass Spectrometer
P1	-	Process 1
P2	-	Process 2
P3	-	Process 3
P4	-	Process 4
P5	-	Process 5
PAHs	-	Polycyclic Aromatics Hydrocarbons
PCA	-	Principal Component Analysis
PNN	-	Probablistic Neural Network
R^2	-	Correlation Coefficient

SC	-	Sensor Centroid
SFE	-	Super Fluid Extraction
SOP	-	Standard Operation Procedure
SSE	-	Sum Squared Error
Std_{min}	-	Minimum Standard Deviation
Std_{max}	-	Maximum Standard Deviation
Str	-	Strategy
SVM	-	Support Vector Machine
SVD	-	Singular Value Decomposition
TH	-	True High
TL	-	True Low
TR1	-	True Region 1
TR2	-	True Region 2
TR3	-	True Region 3
TR4	-	True Region 4
TR5	-	True Region 5
Ω	-	Ohm
Ω_{min}	-	Minimum Ohm
Ω_{max}	-	Maximum Ohm
$Peak_{min}$	-	Minimum Peak Value
$Peak_{max}$	-	Maximum Peak Value

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND

Agarwood is a type of resins which comes from Karas tree mostly found in Asia region. The tree can be obtained directly from forest or from plantations. The Karas tree that is found from forest is actually a controlled plant. The Agarwood source from the controlled plant is actually limited to a certain extent as compared to the tree that is inoculated. In order to increase the Agarwood massively, the Karas tree is planted in well-structured plantations. The harvest of the Karas tree can only be processed after several years until the Agarwood is formed in the core of the tree. The Agarwood is formed whenever the part of the Karas tree is injured. The injury process at the Karas tree occurs via two ways. The first one is through natural process and the other one through inoculation process. The rate of the Agarwood formation through the first process is slower as compared to the second one [1]. Nevertheless, the quality of the Agarwood grading depends on its unique scent, formation and characteristics of the physical of the Agarwood itself [2]. Countries that involves in the Agarwood are clustered into two categories, consumer and producer country. The consumer countries are mainly from Middle East while the producer countries are mostly from South East Asia. It is mainly used for aroma therapy and incents. As the demand on the Agarwood is getting higher, an apparent classification standard for the Agarwood in the Malaysian market will be very useful to consumers and traders in order to control its quality.

1.2 PROBLEM STATEMENT

A lot of works have been done in identifying the Agarwood. The famous method of identification of the Agarwood is through essential oil composition [2-3]. Currently, there are many research which focuses on the quality of the Agarwood

essential oil. In order to achieve the quality of the essential oil, the Agarwood wood must be identified first before being extracted to essential oil. Presently, the grading of the Agarwood quality and oils is classified based on expertise of human sense, color, chemical-based properties and density of the agarwood resin with no scientific accuracy measurement done [3]. However, odor-based approach that relies on experts is still questionable especially in terms of its data measurability and consistency. Based on that issue, this research will use scientific approach using E-Nose which will complement the existing odor-based approach (human experts). However, E-Nose may not possible to be used directly; it needs extra signal processing to process raw data statistically. The data signal processing will increase the possibility of extracting significant features from Agarwood sensor array profile. In this thesis, the intelligent signal processing model will be employed based on E-nose signal in classifying the Agarwood grades and regions. Artificial neural network (ANN) and k-nearest neighbor (k-NN) will be used as the intelligent classifier. Both classifiers have shown great performance and high rate of classification accuracy.

1.3 RESEARCH OBJECTIVE

The main aim of this thesis is to develop an intelligent classification model for the detection of the Agarwood regions and grades using signal processing technique. The proposed classification model must reflect significant achievement in classification performance in terms of classification accuracy and residual errors. The objectives of this thesis are the followings:

1. To pre-process normalized Agarwood measured data using a Principal Component Analysis (PCA) as a pattern analysis technique and a new technique based on the sensor centroid statistical technique to extract Agarwood unique features.
2. To develop an intelligent classification model using k-NN and ANN classifier based on several parameters and extracted Agarwood unique features.
3. To perform comparative study of different k-NN classifiers and ANN models with several strategies and approach in classifying Agarwood.

4. To propose a new data processing technique with a novel sensor centroid features for k-NN classifier.
5. To cross-validate the k-NN classifier and ANN model using several classification performance measures.

1.4 RESEARCH SCOPE

The scope of the thesis include data collection at Forest Research Institute Malaysia (FRIM) research laboratory under controlled condition, based on E-Nose C320 measurements, data analysis using principal component analysis (PCA) and sensor centroid feature extraction technique, classification of Agarwood grades (high and low grade) and regions using selected classifiers. The samples are limited to Malaysian (Melaka, Pagoh, Super Pagoh and Ulu Tembeling) and Indonesian samples. Each data set per sample is ten (10) from each region. The experiment is follows the registered Standard Operation Procedure (SOP) provided by FRIM [4]. The total of overall datasets consists of 10 datasets from each region multiplied by 5 sample regions multiplied by 32 sensor arrays which total of 1600 overall data. The software tool that is used in this work is MATLAB software, and Microsoft excel.

1.5 THESIS LAYOUT

This thesis is organized into six chapters. The content of each chapter is briefed as follows:

Chapter 1 introduces the background of this thesis which includes the Agarwood, its origin and clusters, the type of Agarwood classification factor, the problem being addressed and the objectives and scope of this research.

Chapter 2 reviews the literature related to the work in this thesis. It elaborates the research history in Agarwood and Signal Processing techniques.

Chapter 3 explains the fundamental aspects of data analysis, feature extraction techniques, classification techniques, error analysis, and classification accuracy in this thesis.

Chapter 4 presents the methodology applied in achieving the objectives of this thesis. It covers sample preparation, feature extraction, classification of the k-NN classifier design and the ANN model.

Chapter 5 discusses the data measurements and pre-analysis which include sample preparation, instrumentation setup, data collection, data analysis and feature extraction. The results are represented in various graphical representations. This chapter also discusses the results of Agarwood grades and regions classifications using significant features extracted using k-NN classification techniques and ANN model design. The results are also shown using several significant plots. The classification results of k-NN and ANN classifications techniques are evaluated using several performance measures.

Chapter 6 presents and discusses the conclusion for thesis with some recommended potential future improvements of the current work.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter consists of two main sections; literature reviews on Agarwood (Section 2.2) and the signal processing (Section 2.3). Section 2.2 is the detail reviews on the history of Agarwood grades and regions (Section 2.2.1), sensing and detection (Section 2.2.2), physical-based detection (Section 2.2.3) and chemical-based detection (Section 2.2.4). In Section 2.3, reviews on the chemical-based detection specifically on E-nose (Section 2.3.1), data processing and feature extraction using PCA (Section 2.3.2) and centroid (Section 2.3.3) are presented. Next, general classification technique, Intelligent Classification Technique (Section 2.3.4), classifications analysis from E-nose data (Section 2.3.5) is reviewed respectively. For data preparation, data splitting and randomization (Section 2.3.6) is briefly described. Specific intelligent classification technique is reviewed in K-Nearest Neighbor (k-NN) (Section 2.3.7) and Artificial Neural Network (Section 2.3.8) is reviewed on several applications based on E-nose. The last section of this chapter is reviews on Performance measure that had been done using k-NN and ANN technique (Section 2.3.9). This chapter ends with summary of the overall literature reviews.

2.2 AGARWOOD

Agarwood is a forest product where it is reported as one of the highly valuable commodity [5-6]. Agarwood has different names in different countries. Agarwood is known as Gaharu in Malaysia and Indonesia [2, 7], Jin-koh in Japan [8-9], Ch'en Hsiang in China [8-9] and Oudh in United Arab Emirates [10]. The main markets for

these products are in South and East Asia and the Middle East [11] and it also traded internationally [12].

There are many applications that involve Agarwood in various forms such as in raw Agarwood, processed Agarwood and most commonly burnt into incense especially in cultural activities, religious functions and medicine purpose [8-9]. Normally raw Agarwood is processed into sculpture, beads and essential oil or liquors [8-9], powder and timber pieces [10].

Benefits of Agarwood in medicine which has been found recently are a pharmacological effect on the central nervous system achieved by oral administration or abdominal injection [13-14] and human health [15].

2.2.1 Agarwood Grades and Regions

Agarwood may be classified into various grades: Grade A and B (high grade), C and D (low grade) [16]. While in India they are graded into; Grade I: Black Agar (high grade), Grade II: Bankang and Grade III: Pjutas, Kalaguchi and Grade IV: Dhum (Low Grade) [4]. Grade C of Agarwood is often distilled to obtain Agarwood Oils [16]. Its cost ranges from RM 30,000.00 to RM 40,000.00 per kilogram. Good quality of Agarwood which is widely used as incense [13] and usually requested from a company to produce high quality of Agarwood oil [10].

Another grade of Agarwood is low quality where the Agarwood is touched up with small layer of Agarwood powder and mixed with wax where it is then heated [10] . It costs ranges from RM 10,000.00 to RM 15,000.00 per kilogram while the low quality ones are used for essential oil extraction [2, 10].

Agarwood are graded based on several quality indicators and they are often graded according to physical properties, Agarwood formation , and its unique scent [10, 16], chemical properties [17] and country of origin [10]. Physical properties are often indicated by color, thickness and density [10].

In chemical properties, it has been known by extracting Agarwood into essential oil. Essential oils are volatile compound produced by plants for other purpose besides nutrition. It is categorized under natural product and Agarwood is one of the natural products class. There are several major classes of natural product