WOOD SPECIES RECOGNITION BASED ON PHASE - ONLY CORRELATION (POC) TECHNIQUE

NIK KAMARIAH BINTI NIK ISMAIL

MASTER OF COMPUTER SCIENCE

UNIVERSITI MALAYSIA PAHANG

WOOD SPECIES RECOGNITION BASED ON PHASE - ONLY CORRELATION (POC) TECHNIQUE

NIK KAMARIAH BINTI NIK ISMAIL

Thesis submitted in fulfillment of the requirements

for the award of the degree of

Master of Computer Science

Faculty of Computer Systems & Software Engineering

UNIVERSITI MALAYSIA PAHANG

MAY 2013

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Master of Science (Computer).

Signature

Name of Supervisor	:	PROF. DR. JASNI BINTI MOHAMAD ZAIN
Position	:	DEAN
Date	:	

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged. The thesis has not been accepted for any degree and is not concurrently submitted for award of other degree.

Signature

Name : NIK KAMARIAH BINTI NIK ISMAIL

ID Number : MCC08005

Date :

This effort is dedicated to:

my dearest mother, Nik Nab Binti Nik Hassan

my dearest father, Nik Ismail Bin Nik Hussin

and

my motivation, my family members

ACKNOWLEDGEMENTS

In the name of Allah, invocation and greetings to adoration of Nabi Muhammad S.A.W, thanks to God for giving me strength and patience in finishing my research~~–~~. A sincere gratitude to my supervisor, Prof. Dr. Jasni binti Mohamad Zain for her germinal idea, invaluable guidance, continuous encouragement and constant support in making this research possible. She has always impressed me with her outstanding professional conduct and her belief that a research is only a start of a life-long learning experience. I appreciate her consistent support from my first step as a postgraduate student to these concluding moments. I am truly grateful for her understanding on my situation and condition and her tolerance on my research attitude.

My sincere thanks to all my gratitude to Dr. Tuty Asmawaty Abdul Kadir, Dr.~~Miss~~ Zalili Musa, Dr.~~Miss~~ Rohani Abu Bakar, Miss Noorlin Mohd Ali, Mrs.Nor Amizam Jusoh, Jihan and Che Norehan Che Wahab for their support,idea, knowledge sharing and motivation to complete this thesis.

I acknowledge my sincere indebtedness and gratitude to my parents for their love, dream and sacrifice throughout my life. No appropriate words that could properly describe my appreciation for their devotion, support and faith in my ability to attain my dreams.

Last but not least, thanks to Research Management Center for allocating funding for this research.

ABSTRACT

This research ~~is~~was performed in order to determine ~~the~~ suitable methodology and techniques to ~~indentify the wood species~~ automatically~~.~~ identify wood species. Traditional wood recognition requires ~~skilled~~ employees who are skilled in recognizing wood species. ~~There are~~ The two main problems ~~in this~~ with recognition~~, which are~~ is the lack of skilled employees and ~~the~~ visually impaired employees. This ~~will lead~~leads to ~~the~~ an interruption of the wood recognition process. Besides that, based on ~~the~~ previous ~~researches there are~~research, two techniques ~~applied~~ are used to recognize wood species. These techniques ~~are found that cannot to~~often do not give a high ~~and clear~~ accuracy ~~to the~~ of wood species recognition. ~~According to this~~Accordingly, to improve the accuracy rate ~~to the~~ of wood species recognition, this research ~~is~~was proposed. There are four main steps ~~done~~performed to develop ~~the~~a wood species recognition system~~ which~~. These are; data acquisition, pre-processing, features extraction, and features analysis. The technique used ~~is~~ was the Phase-Only Correlation (POC) technique. The POC technique involves three ~~(3)~~ steps during the feature extraction phase. ~~The~~ These steps are ~~the2D~~the 2D Discrete Fourier Transform (2D DFT), Cross Phase Spectrum, and 2D Inverse Fourier Transform (2D IFT). Each step contains specific ~~calculation~~calculations and ~~algorithm in order~~algorithms to ~~get~~obtain details on ~~woods~~the wood's features and match ~~it~~them with ~~the~~ possible matching species. By applying all ~~the~~ three ~~(3)~~ steps, the ~~testing~~test image is found to ~~be a~~either match or ~~mismatch with~~ not match the training data. ~~Then~~Next, the classification process is ~~done~~performed with the help of the K-Neighbour Network (KNN) technique; where it helps to classify the wood images~~ into its classes.~~. This technique ~~is tested for 8~~was used to test eight wood species~~ which are~~, namely Balau, Ramin, Durian, Melunak, Mersawa Gajah, Rengas, and Sepetir. From ~~the series of~~this testing~~ done~~, it ~~is~~was found that Balau and Ramin wood species ~~are~~were identified with 100% ~~accurately.~~accuracy. Meanwhile, Melunak and Rengas wood species ~~are~~were identified ~~in about~~to approximately 98.5~~%. While for~~%; Mersawa ~~wood species, the accuracy of the testing is about~~ to 95~~%. For the~~ %; and for both Durian and Sepetir wood species~~ obtained~~ , 93.75% accuracy~~.~~ was obtained. Furthermore, for the Mersawa Gajah wood species~~ there are~~, only 91.25% ~~of the data tested wood species are successfully identified. The~~accuracy was obtained. These research results ~~obtained from the research done, it showed that there are~~ show that 96.4% of the wood species tested accurately and were recognized without error.

ABSTRAK

Kajian ini dilakukan bertujuan untuk menentukan metodologi dan teknik yang bersesuaian bagi mengenalpasti spesis kayu secara automatik. Penentuan jenis kayu-kayan masih lagi dibuat berdasarkan pemerhatian mata kasar. Pengenalpastian jenis kayu secara tradisional ini memerlukan pekerja yang benar-benar mahir dalam mengenalpasti jenis kayu. Terdapat dua permasalahan utama dalam kaedah pengkelasan ini, iaitu kekurangan pekerja mahir dan pekerja mengalami masalah penglihatan. Keadaan ini mengakibatkan proses mengenalpasti jenis kayu terganggu. Selain itu, berdasarkan kajian-kajian terdahulu terdapat dua teknik yang telah digunakan bagi mengenalpasti spesis kayu. Teknik- teknik ini didapati tidak dapat memberi kadar ketepatan yang tinggi dan jelas bagi mengenalpasti spesis kayu. Sehubungan dengan itu, bagi meningkatkan lagi kadar ketepatan pengenalpastian spesis kayu kajian ini telah dijalankan. Terdapat empat langkah utama yang dijalankan untuk membangunkan sistem kenalpasti spesis kayu iaitu; perolehan data, pra-pemprosesan data, ekstrak data dan analisis data. Antara teknik yang digunakan adalah teknik ‘phase-only correlation’ (POC). Teknik POC melibatkan tiga langkah semasa fasa ekstrak data. Langkah-langkah ini adalah ‘2D Discrete Fourier Transform’ (2D DFT), ‘Cross Phase Spectrum’, dan ‘2D Inverse Fourier Transform’ (2D IFT). Setiap langkah mengandungi pengiraan dan algoritma khusus untuk mendapatkan maklumat mengenai ciri-ciri kayu dan dipadankan dengan spesies yang mungkin sepadan dengannya. Dengan menggunakan ketiga-tiga langkah, imej yang diuji didapati sama ada sepadan atau tidak sepadan dengan data. Seterusnya, proses klasifikasi dilakukan dengan bantuan K-Neighbour Network (KNN) teknik; di mana ia membantu untuk mengklasifikasikan imej kayu. Teknik ini telah diuji dalam 8 spesis kayu, antaranya ialah Balau, Ramin, Durian, Melunak, Mersawa, Mersawa Gajah, Rengas dan Sepetir. Daripada siri pengujian yang dijalankan,didapati bahawa kayu spesis Balau dan Ramin dapat dikenalpasti 100% dengan tepat. Manakala spesis kayu jenis Melunak dan Rengas dapat dikenalpasti sebanyak 98.5%. Bagi kayu spesis Mersawa jumlah ketepatan yang diuji adalah sebanyak 95% .Kayu spesis Durian dan Sepetir masing-masing memperolehi sebanyak 93.75%. Walau bagaimanapun bagi kayu spesis Mersawa Gajah hanya 91.25% yang didapati dari hasil pengujian yang dijalankan. Justeru, hasil dari kajian secara keseluruhan, di dapati 96.4% dari spesis kayu yang diuji berjaya di kenalpasti tanpa sebarang ralat.

TABLE OF CONTENTS

Formatted: Font: Not Bold

Formatted: Indonesian

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

WBI	Wood Based Industry
MATLAB	Matrix Laboratory
POC	Phase - Only Correlation
KDNK	Keluaran Dalam Negara Kasar
MTIB	Malaysian Timber Industry Board
K-NN	K – Nearest Neighbor
GDP	Gross Domestic Product
MGR	Malaysia Grading Rules

CHAPTER 1

Formatted: English (U.K.)

INTRODUCTION

Formatted: English (U.K.)

Formatted: English (U.K.)

1.1 INTRODUCTION

Tropical rainforests in South~~-~~ East Asia are blessed with more than 15,000 different plant species, of which about 3,000 ~~species~~ can be ~~categorized~~categorised as timber species. Together they made up approximately 33% of the world's timber industry in year 2010 (Forestry Department, 2010). Currently, natural resources such as wood have become scarce and very expensive. Thus, ~~maximizing~~maximising the usage and reducing ~~the~~ rejection are great challenges for the wood industry (Cavalin et al. 2006). In economic history, wood was ~~being~~ conceived as a major factor ~~in contributing largest~~towards contribution of large income to the country.

Based on research ~~from~~by Ariff (2005), the wood-based industry (WBI) had long been an important part of the manufacturing sector, the ~~main~~key driver of economic growth for ~~the country.~~Malaysia. The wood industry has undergone major changes over the years, with downstream activities becoming increasingly important.

In order ~~the~~ to sustain the strong ~~competition,~~competitiveness in~~-~~ highly demanding markets, producers must focus on the quality of their products and increase the efficiency of production process cluster (Baginski et al. 2007). In light of this, various studies have

Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: English (U.K.)
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:

been ~~done~~conducted to develop a system ~~which~~that is suitable and effective ~~to identify~~in identifying the species of wood that issued from time to time.

The three major categories of wood ~~types~~ are softwood, hardwood and composite wood. Softwood is available directly from the wholesale and retail yards or lumber brokers. ~~Whereas for hardwood are~~Hardwood, on the other hand, is used in construction for flooring, architectural woodwork, interior woodwork and panelling. As ~~the~~ for composite wood ~~which~~, it is also known as engineered wood, which includes a range of derivative wood products.

Composite wood is manufactured by binding ~~together~~ the strands, particles, fibres or veneer of wood together with adhesives from composite materials. In the construction industry, choosing and ensuring the right wood species to be used is very important. ~~This is important~~, as different species of wood has different characteristic and features. ~~The~~These differences ~~-~~ might affect the result or products when they are used for different purposes (Tou et al. 2007).

In the wood industry, quality control ~~makes~~is an important aspect to produce a product. Quality control is a procedure or set of procedures intended to ensure that a manufactured product or performed service adheres to a defined set of quality criteria or meets the requirements of the clients or customers. This is because the high quality of product would attract clients or customers interest to purchase and also increase the demand in local or international market. Therefore, wood industry needs an efficient and effective system to ensure the high quality of ~~the product~~products produced.

Wood is valuable as an engineering material ~~–~~ and manufacturing, and in many cases, technology advances have made it even more useful. ~~The~~ One of the advantages of computer vision system is that it would indirectly make a process ~~become~~ more effective ~~where it is~~, as well as consistent and safe for the wood industry product. ~~This will~~ As a result ~~the~~, production of the wood product increased with a ~~highly~~better quality and ~~become potential to~~could potentially compete in the international market. Apart from that, it will

Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: English (U.K.)
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: English (U.K.)
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: English (U.K.)
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: English (U.K.)
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:

also save cost and production time~~ because~~, as it does not require many expert workers. By using computer vision system, accurate result in brief periods can be attended.

At present, humans identify, classify and ~~recognize~~recognise most wood species in order to allow cut-up decision to be made. This operation is ~~labor~~labour intensive and costly (Buehlmann et al. 2007). ~~Recognizing~~Recognising the type of species is an important problem in the wood industry process. Cost reduction in production and inspection process is also a crucial objective for wood manufacturers. This is because the quality of inspection process is manually performed by experts.

Unfortunately, experts cannot ~~recognize~~recognise more than 60 % of the overall category in the wood species. Like other inspection process, it is ~~depended~~dependent on workers' experience until now. ~~The~~As such, the development of a flexible, efficient, reliable, and integrated real~~-~~-time system for industrial application is an essential issue in quality control process for wood manufacturers.

To develop a system for ~~recognizing~~recognising and classifying ~~the~~ wood species, several important aspects should be studied and ~~analyzed.~~analysed. Research based on the nature of the methods and data that form~~ with~~ the images of each type of species to be emphasized is identified. This is to facilitate the determination of the method that will be used in developing the~~ development of~~ system later.

One of available ~~technique~~techniques that can be used to overcome and solve the problem is a computer vision. Computer vision is one of the ultimate unsolved problems in computer science, and solving it, or even small parts of it, creates exciting new possibilities in technology, engineering and even entertainment. Manufacturing sector ~~are~~is growing quickly with computer vision and is now in food processing, pharmaceuticals, wood and paper, plastics, metal fabrication and other industries.

Computer vision is computer imaging where the application does not involve a human to begin in the visual loop. In other words, the images are examined and acted upon

Formatted: Font:
Formatted: Font:
Formatted: English (U.K.)
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: English (U.K.)
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: English (U.K.)
Formatted: Font:
Formatted: Font:
Formatted: Font:

by the computer. Although ~~human~~humans are involved in the development of the system, the final application requires a computer to use the visual information directly (Umbaugh 1998).

The field of computer vision aims to mimic human vision through automated analysis of images and video, using the output from a camera to ~~recognize~~recognise people and objects, and track moving things. Computer vision has applications in robotics, surveillance, remote sensing of the environment, and computer graphics (Bebis et al. 2003).

~~In fact of that, computer~~Computer vision for wood species recognition system ~~in~~ within the wood industry is greatly needed. ~~Thus, wood~~Wood species recognition system ~~was~~has been developed to identify the different species of Malaysian wood. This system, which used ~~capture~~captured image as a sample for characteristic observation~~. Moreover, the wood species recognition system has a~~, has the capability to identify wood species effectively.

1.2 PROBLEM STATEMENT

Traditionally, ~~wood's~~the wood industry used human experts to identify ~~the wood's~~wood species. It takes time to train a person to be competent in identifying the species with a high accuracy. ~~However, by~~Nevertheless, using expert naked eyes to classify ~~it~~ the species have ~~led to~~ several problems, such as ~~inconsistence~~inconsistent recognition results~~, -~~ only workers who are very experienced are able to ~~classify it~~make the classification, and the process of wood species classification is very ~~depended~~dependent on the experts' opinions.

By using the naked ~~eyes~~eye, the process of ~~standardizing~~standardising the species of wood ~~seems to~~can be difficult. One ~~of them is the~~such difficulty is in ~~standardizing~~standardising the decision from ~~different~~ a number of ~~peoples~~people especially in a subjective matter (Musa et al. 2007).

Formatted: Font:
Formatted: English (U.K.)
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: English (U.K.)
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: English (U.K.)
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: Font:
Formatted: English (U.K.)

In addition, manual examination of wood identification can also be ~~very~~ subjective. Besides that, the process of identification ~~which~~that is ~~depending~~dependent on human experts ~~is~~ always ~~having a problem especially~~has issues in managing resources. For example, if ~~a worker is retired,~~the expert retires, the company ~~needs~~would need to make another investment to train a new ~~expert~~ person in identifying the woods species. ~~Moreover,~~Furthermore, such traditional method ~~needs~~requires a ~~huge~~large number of workers and much time to ~~make~~complete a process ~~done~~.

~~Furthermore, the usage of a~~When large ~~labor~~labour force ~~will~~ is used, this leads to increase in cost of production~~. This is because the employer had to pay~~ , as more salaries ~~to many workers. The increasement~~are paid out. Subsequently, the increment in production cost will directly ~~increase~~raise the price of goods in the market and lead to a lower demand from consumers.

~~Currently, there~~There are currently two studies for wood recognition. Both ~~of~~ proposed studies are based on the Grey Level Co-~~occurance~~occurrence Matrices (GLCM). ~~According to the~~ The result of ~~the~~these studies~~, show~~ showed that the accuracy rate for ~~the~~ wood species recognition is not completely successful ~~with~~, as only low accuracy obtained. ~~For the~~ The first study ~~entitled~~, “Wood Species Recognition System” by Bremananth et al. 2009 ~~the study~~, used ~~on~~ 10 Indian woods species, which ~~are~~included, Teak, Ebony, Jack, Oak, Padauk, Sal, Satin, Teak, White Oak and Zebra as the sample. ~~While for~~Meanwhile, the second study ~~entitled~~, “Design of An Intelligent Wood Species Recognition System” by Khalid et al. (2008) used six Malaysian woods species ~~which are Bintangor,~~ Bintangor, Nyatoh, Sesendok, Ramin, Mersawa and Jelutong.

Formatted: Kern at 12 pt

~~Based on those studies, there~~There are several problems found in these studies ~~which in~~, partly due to the GLCM technique. ~~Where the~~ As a result ~~can~~, the outcome could only achieve up to 96 % of accuracy. The other ~~problems are~~ issues include the sensitivity of GLCM ~~is sensitive by~~ towards the change in brightness, and required more filtering techniques in order to remove the noise ~~since~~, as the process in GLCM ~~is~~does not really emphasize ~~on~~ filtering stage.

Since the Phase Only-Correlation (POC) technique has been found successful in ~~most~~many fields of ~~field of the~~ studies ~~done~~, especially in face recognition system ~~studies~~, the researcher has decided to apply POC in order to ~~recognize~~recognise the wood species. ~~By then~~In addition, there are certain techniques applied to enhance POC for a better result.

1.3 OBJECTIVES

The main objective of this research is to propose ~~the~~ a wood species recognition technique. This objective can be divided ~~as follows~~into the following:

i) To develop wood species recognition based on phase only-correlation (POC) technique.

ii) To classifies wood species using K-~~-~~Nearest ~~Neighbor~~Neighbour technique.

iii) To evaluate and compare the proposed technique ~~with~~against the GLCM ~~technique~~method.

1.4 RESEARCH SCOPE AND APPROACH

This research is focused on the development of an automated recognition and classification system, which consists of image acquisition, image enhancement, image segmentation, feature extraction and classification. The entire recognition and classification system is developed in MATLAB. The scopes of this research can be divided into four (4), they are:

i) This project is developed to ~~recognize~~recognise and classify the wood species by using ~~grayscale~~grey scale image of 450 x 300 sizes.

ii) This project has been developed using MATLAB as a data-manipulation software package that allows data to be ~~analyzed~~analysed and ~~visualized~~visualised using existing functions and used-designed ~~programs~~programmes.

iii) This project has used ~~on~~ eight (8) Malaysian woods species, which are Balau, Ramin, Durian, Melunak, Mersawa, Mersawa Gajah, Rengas and Sepetir.

iv) ~~Numbers~~The number of training and testing data ~~are~~is eighty (80) images for each species.

1.5 DISSERTATION OUTLINE

This thesis is divided into five (5) Chapters and ~~organized~~organised as ~~follows.~~followed. Chapter 1 introduces the background of the wood industry, the problem statement, the objectives and the research scope and approach. ~~The~~Chapter 2 presents the literature review on wood species classification and methods used to accomplish it, ~~give~~as well as provide knowledge on wood identification and techniques that are previously used in wood species identification~~ is presented in Chapter 2.~~, The detail of methodology, testing and prototype of research development is presented in Chapter 3~~.~~ while Chapter 4 ~~will be discussed~~discusses the result obtained and ~~evaluate~~evaluates the proposed techniques. Finally, the summary of the thesis and conclusion to the work is presented in Chapter 5.

Formatted: English (U.K.)

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

Malaysia, located in South~~-~~ East Asia, is a federation of states. It spreads across two separate geographical regions~~ they ~~, which are Peninsular Malaysia and Borneo Island. ~~Tropical~~The tropical rainforests in ~~south ~~South East Asia are blessed with more than 15,000 different plant species, of which about 3,000 species can be ~~categorized~~categorised as timber species.~~ The~~ Malaysia's timber industry continues to be a major export earner for the nation. In 2009, Malaysia exported a total of RM 19.4 billion worth of timber products.

In the same year, the country's import ~~for~~of timber product, excluding furniture, amounted to RM 1.1 billion. This indicates that Malaysia is not only a major exporter of timber products but is also becoming a significant importer to supply its domestic timber materials (Source: Malaysia Timber Industry Board, 2009).

The growth ~~was also~~has been attributed to the structural transformation of the timber industry, shifting from primary to secondary processing and complemented by the development of synergetic industrial clusters. In line with the industry's encouraging ~~the~~ growth~~ of the industry~~, the Federal Government has made an allocation of RM1 billion for a 15-year ~~program~~programme to plant 375,000 ~~hector ~~hectares of land with high~~-~~-value timber trees by 2020.

Nowadays, wood species recognition is widely used in various areas and environments. For ~~examples~~example, it is used in the construction industry and manufacturing ~~industry.~~sector. Wood species recognition ~~is an implementation on identifying the~~ refers to the identification of different species of wood ~~with~~through the capture of wood sample or the characteristics observed.

In the wood industry, wood ~~material~~materials must be examined before ~~it is~~ they are selected to produce a product. This is important because ~~the~~ different species of wood has ~~the different~~varied characteristics and features, ~~these~~which might affect the result or products if they are used for different purposes.

~~According to (Tou et al., 2007)~~ According to Tou et al. (2007) in their study, Computer Vision-based Wood Recognition System has stated that in the construction industry, choosing and ensuring the right wood to be used is very important. Similarly, the manufacturing of wood products such as cupboards, tables and chairs must also ~~be using~~use wood material of a certain quality. If the wood materials ~~which~~that are not strong enough are used at critical areas such as the roof truss, part of the house may collapse after a period of time (Lew, 2005).

The safety issue of these products is important, since collapsing ~~house and~~houses or chairs that suddenly break might cause serious injuries or even ~~led~~lead to fatal results~~, therefore~~. As such, the types of wood used must be properly chosen and verified.

Traditionally, wood recognition ~~has uses~~used human expert to identify the wood's species. It takes a long time to train a person to be competent in wood identification and to identify wood species with high accuracy. Furthermore, manual examination of wood identification can be very subjective.

Besides that, the process of identification ~~which~~that is ~~depending~~dependent on human expert ~~are always~~ tends to have ~~a problem~~issues, especially in managing resources. For example, if a worker ~~is retired,~~retires, the company needs to make another investment to train a new expert to identify ~~the~~ woods species. Moreover, traditional method needs a huge number of workers and much time to make a process done.

In this chapter, several previous studies have been reviewed in order to ~~give~~provide knowledge about the important aspects of developing a system for wood species recognition. Several aspects must be concerned in obtaining the more effective system, such as methods, techniques and algorithms that are suitable to ~~implement~~be implemented in wood recognition system.

2.2 CLASSIFICATION OF MALAYSIAN WOOD

In Malaysia, most of the timbers produced are hardwoods. From 3,000 species of ~~tree~~trees in the forest, only 12 species are softwoods. Of the 3,000 species, a total of 677 species can achieve girth of 1.2m at breast height. ~~There~~These are the species that are being exploited for timber. The timbers that are known in the international market, ~~totaling~~totalling 480 species, have been introduced through the Malaysia Grading Rules (MGR) (Menon, 2004).

In the 1984 edition of the MGR, the timbers are divided into 100 ~~timbers~~timber groups ~~comprising~~that comprised the commercial timbers in Malaysia, and these ~~include~~included the softwood timber group. Generally timber can be classified into:

(i) Heavy hardwood (HHW)
(ii) Medium hardwood (MHW)
(iii) Light hardwood (LHW)
(iv) Softwood

The ~~details~~detailed features such as density, uses, and durability of each classification can be referred to Table 2.1; while Table 2.2 shows the species for each classification of ~~Malaysia~~Malaysian timber.

Table 2.1: The features of wood classification

Formatted: English (U.K.)

Groups	Heavy Hard Woods	Medium Hardwood	Light Hardwood	Softwood
Density	Possess density above 880 kg $/m^3$ at 15% moisture content and are naturally durable	Moderately heavy ~~-~~and range in density from about 720 to 880 kg $/m^3$ 15% moisture content	Below 270 kg $/m^3$ at 15 % moisture content	Range from about 385 to 735 kg $/m^3$
Uses	Construction timbers.	Construction purpose.	Furniture, decorative ~~paneling, ect.~~panelling, etc.	Mainly as decorative plywood and ~~paneling.~~panelling.
Durability	The durability of the timbers is due to the fact that most of them contain some toxic materials within their tissues such as alkaloids and other substances abhorrent to wood destroying agents. Such timbers can be used in most exposed conditions without undergoing preservative treatment. However, the sapwood of these timbers requires treatment as it is not naturally durable.	Moderately durable timbers. Some timbers in this class, such as kempas and tualang, are heavy and strong, but under tropical conditions, their natural durability is insufficient for use in exposed condition and in ground contact unless they are properly treated with chemical preservatives. Under temperate conditions most of the timbers in this group are naturally durable as the climatic conditions are likely to be less conducive to the activity of wood destroying agents.	The timbers are not naturally durable in tropical climate but ~~are~~ generally quite durable in temperate conditions. The timbers in this group require preservative treatment as precaution agents against wood destroying agents, such as fungi and insects.	Softwood timbers are differentiated from hardwood timbers by the presence ~~by the~~of tracheids instead of vessels (pores). The three (3) groups of softwood species are dammar minyak, podo and sempilor. However, only dammar minyak is of commercial use at the moment.
Species	14 species groups (See Table 2.2)	36 species groups (See Table 2.2)	47 species groups (See Table 2.2)	3 species groups (See Table 2.2)

Source: Menon (2004)

Table 2.2: Species of each classification of Malaysia timbers

Heavy Hardwood	Medium Hardwood	Light Hardwood		Softwood
1. Balau 2. Red Balau 3. Belian 4. Bitis 5. Chengal 6. Giam 7. Kekatong 8. Keranji 9. Malagangai 10. Merbau 11. Penaga 12. Penyau 13. Resak 14. Tembusu	1. Alan Batu 2. Bekak 3. Derum 4. Entapuloh 5. Geriting/teruntum 6. Kandis 7. Kapur 8. Kasai 9. Kayu Malam 10. Kedang Belum/Tulang Daing 11. Kelat 12. Keledang 13. Kempas 14. Keruing 15. Keruntum 16. Kulim 17. Mata Ulat 18. Mempening 19. Mengkulang 20. Meransi 21. Merawan 22. Merbatu 23. Merpauh 24. Mertas 25. Nyalin (Minyak Berok) 26. Pauh Kijang 27. Perah 28. Petaling 29. Punah 30. Ranggu 31. Rengas 32. Semayur 33. Senumpul 34. Simpoh 35. Tampoi 36. Tualang	1. Alan Bunga 2. Ara 3. Babai 4. Bayur 5. Berangan 6. Bintangor 7. Binuang 8. Dedali 9. Durian 10. Geronggang 11. Gerutu 12. Jelutong 13. Jongkong 14. Kedondong 15. Kelumpang 16. Kembang semangkok 17. Ketapang 18. Kungkur 19. Laran 20. Machang 21. Mahang 22. Medang 23. Melantai 24. Melunak 25. Mempisang 26. Meranti Bakau 27. Meranti,Dark Red 28. Meranti,Light Red 29. Meranti, White 30. Meranti, Yellow 31. Merbulan 32. Mersawa 33. Nyatoh 34. Pelajau 35. Penarahan 36. Perupok 37. Petai 38. Pulai 39. Ramin	40. Rubberwood 41. Sengkuang 42. Sentang 43. Sepetir 44. Sesendok 45. Terap 46. Terentang 47. White Seraya	1. Damar Minyak 2. Podo 3. Sempilor

Source: Menon (2004)

2.3 GENERAL STRUCTURE OF WOOD

Wood is formed as a result of continuous secondary growth in cells. The complicated nature of its structure will be more readily understood by knowing how wood is formed in a growing tree. Figure 2.1 ~~illustrated~~illustrates the structure of wood, which consists of sapwood, heartwood, pith, outer bark, inner bark and cambium elements. The description of each element is stated in Table 2.3.

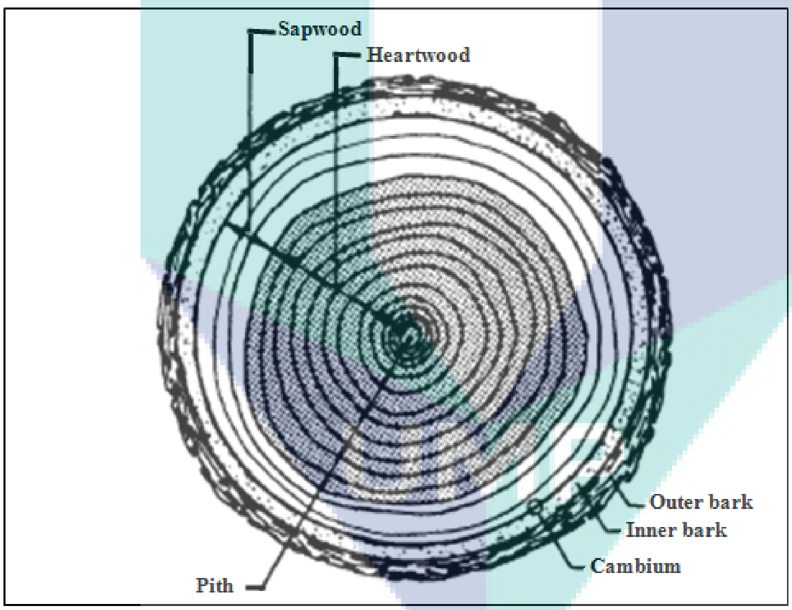


Figure 2.1: The growth rings and the relative positions of pith, sapwood, heartwood, cambium, and bark

Source: Arntzen & Charles (1994)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Table 2.3: Element of wood structure

Formatted: English (U.K.)

Type	Description
Pith	When the seed of a tree germinates, it forms a shoot consisting of thin medulla of spongy tissue.
Cambium	The pith completely enclosed within a thin meristematic tissue
Bark	Which in turn is protected on the outside by ~~a~~ tissues
Inner Bark	Produces new cells of a different kind giving rise to the formation of new phloem or inner bark.
Heartwood and Sapwood	The ~~function~~functions are : a. Conduction of water and dissolved mineral salt solutions from the roots to the leaves for the manufacture of plant food materials. b. Storage and distribution of manufactured and reserved food materials. c. Provision of mechanical strength to the tree as a whole.

Source: Menon (2004)

Figure 2.2 illustrates the three-dimensional differences between hardwood and softwood cell structures. The similarities between hardwoods and softwood are in several structural elements such as ray cells, early wood and latewood. Hardwood structure, nevertheless, is more complex than softwood structure, and varies considerably between species.

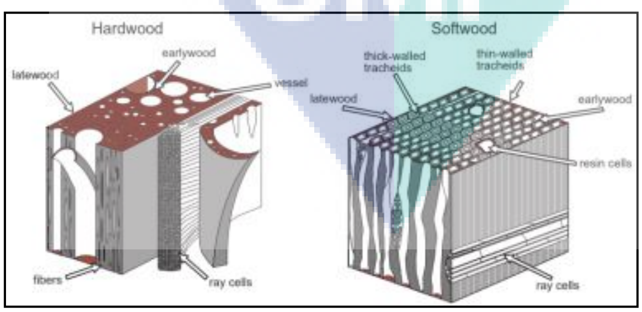


Figure 2.2: Three-dimensional cell level comparison between hardwood and ~~softwoods~~softwood

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Source: Bond et al. (2011)

The majority of hardwood volume is composed of ~~fiber~~fibre cells that offer structural support to the stem. The major difference between hardwood and softwood is the presence of ~~fiber~~fibre and vessel elements that exist in hardwoods only. The main function of vessel elements is water conduction. Vessel elements can vary greatly in size, number and spacing from one species to another, and from earlywood to latewood. The ~~detail~~detailed description of hardwood and softwood are illustrated in Table 2.4.

Common in manual processes, the experts ~~will~~would combine the structure features together with physical features in order to identify the species of woods. The most important physical features used in wood identification are ~~color~~colour, weight, hardness, texture, grain, Figure and smell. The ~~detail~~detailed descriptions of each physical feature are described in Table 2.5.

Table 2.4~~:~~ Description of softwood and hardwood elements

Element	Features	
	Softwood	Hardwood
Vessels	Not ~~exist~~existence	The transverse section of wood ~~are seem~~is seen as small, round or oval holes in the wood. On a longitudinal surface, they appear as small grooves or scratches, running parallel to the axis of the stem.
Thick-walled Tracheids	~~For more~~More than 90% of softwood elements~~,~~ are long, tube-like cells with closed ends, numerous holes (pits) on the side walls.	Not ~~exist~~existence
Early wood	Cells have large diameters and thin cell walls.	Within each growth ring, a band of large earlywood vessels (pores) is clearly visible to the naked eye
Latewood	Thick - walled, smaller diameter cells	A band of latewood vessels appears much smaller and requires the use of a hand lens to see.
Rays	Similar to those of hardwoods but due to lack of pronounced variation in ray structure between woods of different softwoods, their value, in differentiating between different kinds of wood, is almost nil.	Ribbons or plate shaped strands of parenchymatous tissue aligning horizontally in wood.
~~Fibers~~Fibres	Not ~~exist~~existence	The main function of the tissue is to provide mechanical strength and rigidity to the wood.
Resin cells	Presence or absence of resin canals.	Not ~~exist~~existence
Other Structure Features		
Parenchyma	Similar to the hardwoods except that the cells are comparatively sparse.	Storage and distribution of reserve food materials. It consist of two distinct ~~system~~systems, which are: i) Vertical system known as wood parenchyma ii) Horizontal or radial system known as ray or parenchyma or as ray

Source: Menon (2004)

~~Commonly~~

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Table 2.5: Physical features of woods

Formatted: English (U.K.)

Physical Features	Description
~~Color~~Colour	Each wood has a characteristic ~~color~~colour of its own and this helps identification to some extent.
Weight	Made up of chemical compounds, which include cellulose, hemicelluloses and lignin.
Hardness	~~Define~~Defined as the degree of resistance of wood to cutting, shaping, nailing, abrasion, etc.
Texture	The term refers to the quality of wood, which is determined by the relative sizes of the basic structure elements and their arrangement.
Grain	Often wrongly applied to refer to the texture of wood.
Figure	Refers to the attractive designs in wood produced by the arrangement of ~~the~~ elements or by uneven ~~coloring.~~colouring.
Smell	A few woods have a smell characteristic ~~smell~~, which can be used in identification. Some of these timbers are : a. Keruing – resinous smell b. Kapur – camphory smell c. Kulim – garlic - like smell

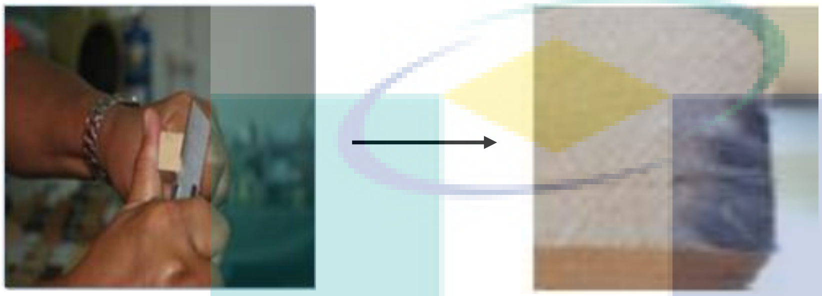
Source: Menon (2004)

2.4 MANUAL EXAMINATION PROCESS OF WOOD IDENTIFICATION

Timber is generally examined in two ways, namely with the naked eye and then, with the aid of a magnifier. First is the examination of wood species with the naked eye. This examination uses the senses without the aid of any accessories. During this process, the examiner needs to consider both structure and physical elements of wood before naming the wood species. The structure and physical ~~element~~elements are already stated in the previous section and ~~illustrate~~illustrated in Table 2.4.

The second way to examine the wood species is using a magnifier. ~~The~~This examination consists of three (3) steps, which require equipment and procedures as illustrated in Figure 2.3 to Figure 2.5.

Step 1



(a) A sharp pocket knife is used to process.
make a cross~~-~~-section of wood.

(b) The wood after cross~~-~~-section

Figure 2.3: Step 1 ~~for~~ of the process ~~of~~ in identifying the wood sample by using a sharp pocket knife

The equipment required for the purpose of identification process is a sharp pocket knife and a hand lens of 10x magnification.

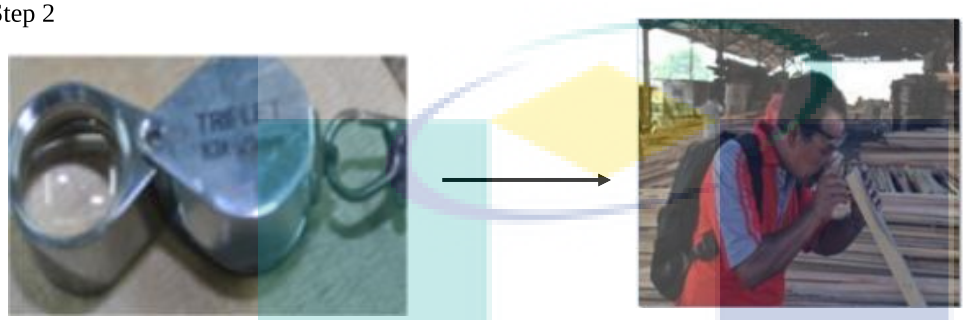
The first step is ~~using~~to use the knife (Figure 2.3). A sharp blade is required for good surface preparation. If the knife is not sharp, the section prepared will be rough and the finer details of the structure may be disturbed. A razor blade is best ~~to~~for use.

A good clean surface is when cells are cleanly cut rather than torn. ~~It~~ Nevertheless, it should ~~also~~ not ~~to~~be cut ~~it~~ too deep~~. Deep~~ as deep cuts ~~will~~would resort in torn ~~fibers~~fibres in the wood section and possible injury to examiner's hands and fingers. Wood samples should preferably be cut in such a way ~~as to expose~~that it exposes the true transverse (cross~~-~~section), radial and tangential surfaces.

On the end surface (transverse section), which should be truly horizontal to the grain then, a clean cut is made~~. This is done~~ by a single movement of the knife. Wetting the surface with water can be helpful in getting a good, clean section.

Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)

Step 2



(a) Hand lens of ~~10 x~~10x magnification used using
to examine the wood structure. magnification.

(b) Process of identifying the wood
a hand lens ~~10 x magnifications~~of 10x

Figure 2.4: Step 2 ~~for~~ of the process ~~of~~ in identifying the wood sample by using hand lens

The second (Figure 2.4) step involves the use of the hand lens. To view the cellular features of wood, it is very useful to have a 10x magnification hand lens to magnify the wood section. The most commonly used lens for identification is the folding type.

The transverse section is then examined by the aid of the hand lens in the way described above (Figure 2.4). If the initial examination shows that the features of the wood are not well defined, a little moisture is applied to the section with a moist finger. This, in most cases, improves the visibility of the wood features.

Formatted: English (U.K.)

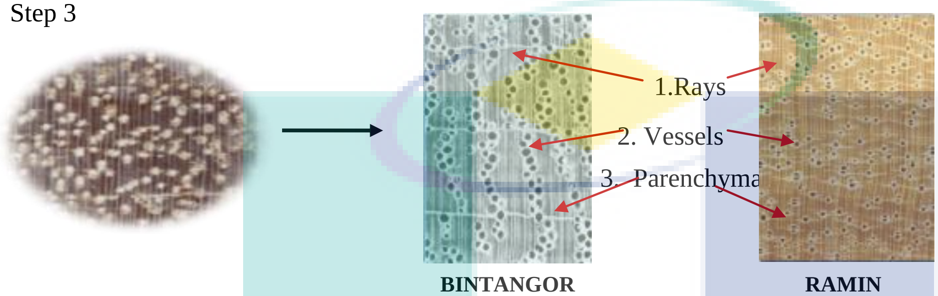
Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Step 3



(a) Structure of wood identified by the hand lens

(b) Three (3) key elements of wood structure considered

Figure 2.5: Step 3 for process of identifying the wood sample

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Step 3 (Figure 2.5) engages the structure features considered to examine the wood sample. The elements of wood structure considered are rays, vessels and parenchyma. The details can be obtained from ~~the~~ Table 2.3. By using a hand lens, the examiner can determine if vessel elements are present or not, thus separating hardwoods from softwoods.

In order to determine the sample of woods, some characteristics are considered~~.~~ ~~Those characteristics include~~, including sapwood, heartwood, ~~luster~~lustre, strip Figure, ring Figure, texture, grain and air dry density. Each characteristic distinguishes one wood species.

For example, to identify the Balau species, the general characteristic of its ~~luster~~lustre is considered. The ~~luster~~lustre of the Balau species should be medium and not too high. ~~In the meanwhile~~Meanwhile, the heartwood of Ramin species should be cream-white to pale straw in colour in order to distinguish it.

Other than the general characteristics, there are structures and other features of wood species that can be considered~~ in order~~ to examine the wood species. The structures are growth ring, parenchyma, rays, ripple marks and intercellular. Table 2.6 and Table 2.7 ~~give~~provide details on general characteristics of wood species and the structure and other features found in the wood species~~ accordingly~~, which are usually used in determining the sample of woods.

Table 2.6: General characteristics of wood species

General Characteristics	Wood Species							
	Balau	**Ramin**	**Rengas**	**Melunak**	**Mersawa**	**Sepetir**	**Durian**	**Mersawa Gajah**
Sapwood	Well defined	Not differential	Well defined , Pink-brown	Not well-defined, lighter-coloured	Not differential	Distinct Light- pink	Lighter, Coloured, Distinct	Not well-defined, Lighter- coloured
Heartwood	Yellow and red brown, brown	Cream-white to pale-straw	Blood-red, streaky	Pink-brown, or red-brown	Yellow-brown to straw-brown	Brown, gold and red brown	Pink-brown, red-brown	Yellow-brown to straw-brown
Luster	Medium	Not particularly	High	-	-	Not significant	-	-
Strip Figure	Subtle	-	-	Radial surface	Subtle	Radial surface	-	Subtle
Ring Figure	-	-	-	-	-	Tangential surface	-	-
Texture	Fine and even	Fine and even	Variable	Fine and even	Coarse and even	Fine and even	Coarse and uneven	Coarse and even
Grain	Interlocked	Shallowly interlocked	Interlocked	Interlocked	Interlocked	Shallowly interlocked	Slightly Interlocked	Interlocked
Air dry density	Over 800 kg/m^3 to < 1155 kg/m^3	Average of 640 kg/m^3	From 672 to 992 kg/m^3	Average of 656 kg/m^3	Average of 640 kg/m^3	Average of 672 kg/m^3	Average of 688 kg/m^3	Average of 640 kg/m^3

Source: Menon (2004)

Table 2.7: Structure and other features of wood species

Structure	Wood Species							
	Balau	**Ramin**	**Rengas**	**Melunak**	**Mersawa**	**Sepetir**	**Durian**	**Mersawa Gajah**
Growth ring	-	-	Concentric	Concentric	-	Distinct	-	-
Vessels	Simple, medium-sized, few, filled with tyloses, no deposits	Simple, medium-sized, few, no tyloses and deposits	Simple, medium- sized, few, tyloses common no deposits	Simple, medium-sized, few or many, tyloses, deposits present	Simple, medium-sized, few ,no tyloses and deposits present	Simple, medium-sized, few ,no tyloses, deposits common	Simple, moderately large, few ,no tyloses and gum like deposits	Simple, medium-sized, few ,no tyloses and deposits present
Parenchyma	Apotracheal, paratracheal	Paratracheal	Apotracheal, paratracheal	Apotracheal, paratracheal	Apotracheal, paratracheal	Vasicentric, Apotracheal	Paratracheal	Apotracheal, paratracheal
Rays	Fine	Fine	Fine	Fine	Two sizes	Fine	Two sizes	Two sizes
Ripple marks	-	-	-	Distinct	-	-	Not always visible	-
Intercellular canals	Small, concentric formation, filled with white resin	-	Horizontal, colourless sap, black blotches on longitudinal surfaces	-	Vertical, smaller to as large as the vessels, resin canals in tangential series	Vertical, smaller than vessels	Traumatic type filled with gummy substances rarely present	Vertical, smaller to as large as the vessels, resin canals in tangential series
Other features								
Burning splinter test	Splinter burns to ash	Splinter burns to ash	Splinter burns to charcoal	Splinter burns to ash	Splinter burns to ash	Splinter burns to ash	Splinter burns to ash	Splinter burns to ash
Froth test	Negative	Negative	Negative	Negative	Negative	Negative	Negative	Negative

Source: Menon (2004)

2.5 PATTERN RECOGNITION

Recognition is regarded as a basic attribute of human beings, as well as other living organism. A pattern is the description of an object. ~~A~~The human being is a very sophisticated information system, partly because he posses a superior pattern recognition capability. ~~According to the~~The nature of patterns recognition can be divided into two major types (Tou & Gonzalez, 1974)

i. Recognition of concrete items

This may be referred as the sensory recognition, which includes visual and aural pattern recognition. This recognition process involves the identification and classification of spatial and temporal patterns. Examples of spatial patterns are characters, fingerprints, physical objects~~,~~ and image. Temporal patterns include speech waveforms, time series, electrocardiograms~~,~~ and target signatures.

ii. Recognition of abstract items

On the other hand, an old argument~~,~~ or a solution to a problem can be ~~recognized.~~recognised. This process involves the recognition of abstract items and can be termed conceptual recognition. Recognition of concrete patterns by human beings may be considered as a physical stimulus.

Human recognition is in ~~a~~ reality a question of estimating the relative odds that the input data can be associated with one of a set of known statistical populations, which depend on our past experience and which form the clues, and the prior information for recognition.

Thus, the problem of the pattern recognition may be regarded as one of the discriminating ~~the~~ input data between populations via the search for features or invariant attributes among members of a population.

A pair containing an observation and a meaning is namely a pattern. ~~Inferring~~Inferred meaning from observation is pattern recognition. Designing a pattern recognition system is establishing a mapping from measurement space into the space of potential meanings, whereby the different meanings are represented in this space as discrete target points~~.~~ (Zheng & He, 2005).

In general, computerised pattern ~~computerized~~ is a complex procedure ~~which~~that requires varieties of techniques that successively transform directly the iconic data into usable information for recognition system. During the past 30 years, pattern recognition was inversed greatly. The ~~needs~~need for theoretical methods and experimental software and hardware are increasing rapidly.

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Nowadays, applications for pattern recognition ~~was~~are used in many areas such as character recognition, target detection, medical diagnosis, biomedical signal and image analysis, remote sensing, identification of human faces and of fingerprints, reliability analyses, socioeconomics, archaeology, speech recognition and understanding, machine part recognition, automatic inspection, and etc.

Formatted: English (U.K.)

Formatted: English (U.K.)

Recently, many methods of pattern recognition have been proposed. Those methods in general are not only used for objects in visual images but also applicable to other types of real world entity. Pattern recognition is often achieved using linear and quadratic discriminates (Fisher, 1936), the k-nearest neighbour classifier (Hodge & Austin, 2005) or the Parzen density estimator (Girolami, 2003), template matching (Meijer, 1992) and Neural Networks (Hush & Horne, 1993).

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Watanabe~~,~~ (1985) said it is an entity, vaguely defined, that could be given a name e.g. fingerprints and handwritten word. Pattern recognition techniques are used to classify physical objects (2D or 3D) or abstract ~~multidimensional~~multi-dimensional patterns known or unknown categories. Pattern recognition can be designed using the following main approaches ~~which are~~ Template Matching, Statistical methods, Syntactic methods and Neural networks (Shinde & Deshmukh 2011).

Table 2.8 shows ~~the approach of~~ pattern recognition according to its approach, representation, recognition function and typical criterion.

Table 2.8: Approach in Pattern Recognition

Formatted: English (U.K.)

Formatted: English (U.K.)

Approach	Representation	Recognition Function	Typical Criterion
Template Matching	Samples, pixels, curves	Correlation, distance measure	Classification error
Statistical	Features	Discriminant function	Classification error
Syntactic or Structural	Primitives	Rules, grammar	Acceptance error
Neural Network	Samples, pixels, features	Network Function	Mean square error

Source: Shinde & Deshmukh (2011), Jain et al., (2000)

~~The interest in the area of~~Interest in pattern recognition has been renewed recently due to emerging applications, which are not only challenging but also computationally more demanding (see Table 2.9). These applications include data mining (identifying a pattern e.g., correlation, or an outlier in millions of multidimensional patterns), document classification (efficiently searching text documents), financial forecasting, ~~organization~~organisation and retrieval of multimedia databases, and biometrics (personal identification based on various physical attributes, such as face and fingerprints).

Jain et al. (2000) has identified a novel application of pattern recognition~~,~~ called affective computing, which ~~will give~~gives a computer the ability to ~~recognize~~recognise and express emotions, to respond intelligently to human ~~emotion~~emotions, and to employ mechanisms of emotion that contribute to rational decision –making.

~~A~~The common characteristic of a number of these applications is that the available features (typically~~,~~ in the thousands) are not usually suggested by domain experts~~,~~ but must be extracted and ~~optimized~~optimised by data-driven procedures.

The rapidly growing and available computing power, while enabling faster processing of huge data sets, has also facilitated the use of elaborate and diverse methods for data analysis and classification. At the same time, demands ~~on~~for automatic pattern recognition systems ~~are~~is rising enormously due to the availability of large databases and stringent performance requirements such as speed, accuracy, and cost. In most existing ~~research~~researches on application ~~for~~ of pattern recognition, they used a combination of multiple ~~approach~~approaches and methods to achieve optimal ~~result~~results (Jain et al., 2000).

Pattern recognition is applied in many areas such as machine learning, statistics, mathematics, computer science, biology and others. There are many problems that appeared while designing the application for pattern recognition. Many of these problems can indeed be solved. In order to solve these problems, more complex learning, searching and ~~optimization~~optimisation algorithm are developed (Bezdek & Pal, 1992). Table 2.9 shows the details ~~for~~of examples of pattern recognition applications comprising problem area, applications, input pattern and pattern classes.

Formatted: Caption, Indent: First line: 1.25 cm

Formatted: Font: Bold

Table 2.9: Examples of Pattern Recognition Applications

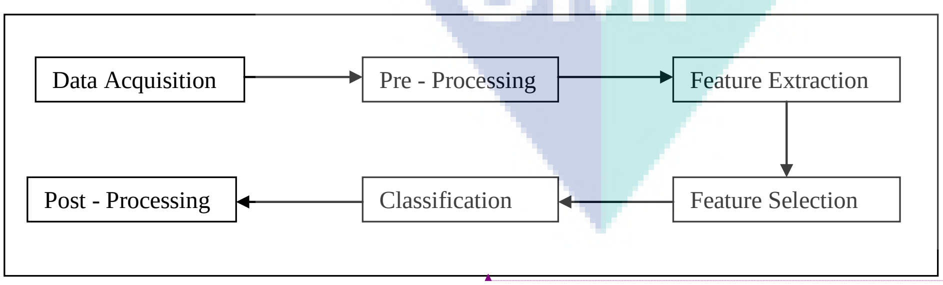
Problem Area	Applications	Input Pattern	Pattern Classes
Bioinformatics	Sequence Analysis	DNA/Protein sequence	Known types of genes or pattern
Data Mining	Searching for meaningful patterns	Points in multidimensional space	Compact and well separated clusters
Document classification	Internet search	Text document	Semantic categories
Document Image Analysis	Optical character recognition	Document image	Alphanumeric characters, word
Industrial Automation	Printed circuit board inspection	Intensity or range image	Defective/non-defective nature of product
Multimedia Database Retrieval	Internet search	Video clip	Video genres
Biometric Recognition	Personal identification	Face, iris, fingerprint	Authorized users for access control
Remote Sensing	Forecasting crop yield	Multi spectral image	Land use categories, growth patterns of crop

Speech Recognition	Telephone directory	Speech waveform	Spoken words
Medical	Computer aided diagnosis	Microscopic image	
Military	Automatic target recognition	Optical or infrared image	Target type
Natural Language Processing	Information extraction	Sentences	Parts of speech

Source: Jain et al. (2000)

2.6 GENERAL PROCESS OF PATTERN RECOGNITION

In general process of pattern recognition can be divided into several ~~phase.~~phases. As shown in Figure 2.6~~. It~~, it starts from data acquisition, pre-processing, feature extraction, feature selection, classification and post-processing.



Formatted: English (U.K.)

Figure 2.6: General process of pattern recognition

Formatted: English (U.K.)

Source: Shinde & Deshmukh (2011)

2.6.1 Data Acquisition

~~First~~The first phase in general process pattern recognition is data acquisition. During this phase, data collection is performed~~. Which~~, of which image in specific format such as JPEG, BMT, GIF, PNG, TIF etc. ~~are~~is required as an input image. This image may be acquired through scanner, digital camera or any other suitable digital input device (Pradeep, 2010).

In order to process the input image, ~~once~~one should consider ~~measure~~measuring physical variable such as bandwidth, resolution, sensitivity, distortion, latency and others (Jain et al., 2000). Before, a pattern vector is made up of a set of measurements, these ~~measurement~~measurements need to be performed using some technical equipment and be converted into numerical form (Lampinen, 1997).

2.6.2 Pre - Processing

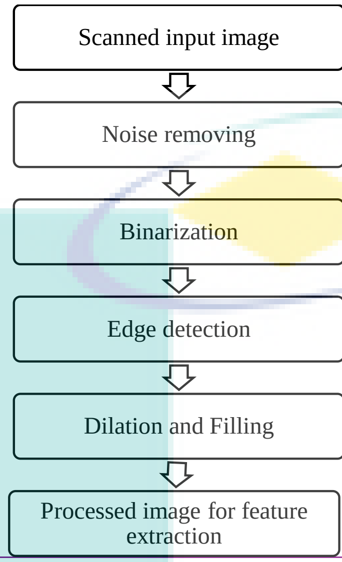
Pre-processing phase will enhance a document image, in order to prepare it for the next phase in a pattern recognition system. Pre-processing phase needs colour, grey-level or binary document images containing text or graphics to be processed. In pattern recognition systems, most of the applications use grey or binary images since processing colour images is computationally high (Alginahi, 2010). Thus, pre-processing phase usually involves noise filtering, smoothing and ~~normalization~~normalisation (Jain et al., 2000), (Shinde & Deshmukh, 2011), (Zheng & He, 2005).

The various tasks performed on the image in ~~preprocessing~~pre-processing phase are shown in Figure 2.7. ~~The pre~~Pre-processing is a series of operations performed on the scanned input image. It essentially enhances the input image to ease the next phase of pattern recognition. Noise removal process is aimed to improve the quality of input image.

Noise, such as salt- and ~~peper~~pepper noise, in ~~image are~~images is usually removed by Average filtering technique, Minimum filtering technique and Maximum filtering techniques. Binarization process converts a ~~gray~~grey scale image into a binary image using global thresholding technique (Nomura et al., 2009).

Detection of edges in the binarized image using the sobel technique, dilation of the image and filling of the holes present in it, are the operations performed in the last two stages to produce the ~~preprocessed~~pre-processed image (Gonzalez et al., 2004) suitable for segmentation (Pradeep, 2010) (Pan, 2005).

Formatted: Indent: First line: 0 cm



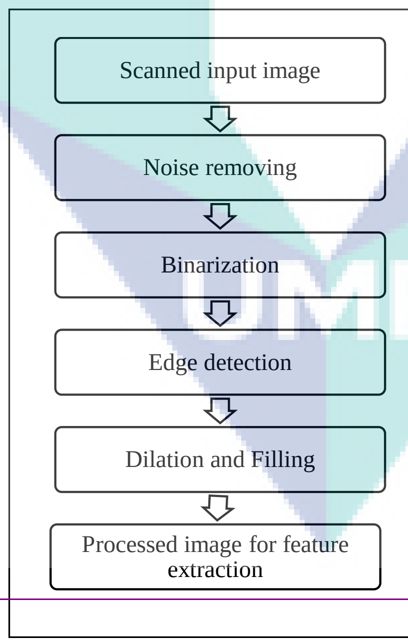


Figure 2.7: Processing Stage

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Source: Pradeep (2010)

Formatted: English (U.K.)

2.6.3 Feature Extraction

Feature extraction is one of the important steps in pattern recognition. It extracts a set of descriptors, various characteristics attributes, the relevant information associated to form a representation of input pattern (Ashoka, 2012). It also finds a new representation in terms of features (Shinde & Deshmukh, 2011).

Formatted: Indent: First line: 0 cm

Features should be easily computed, robust, insensitive to a variety of distortions, and variations in the images. Two kinds of features are used in pattern recognition problems. The first one has clear physical meaning, such as geometric and statistical. Another kind of features has no physical meaning and it is called the mapping features (Zheng & He, 2005).

In feature extraction, most methods are supervised. These approaches need some prior knowledge and ~~labeled~~labelled training samples. There are two kinds of supervised methods ~~are~~ used, which are Linear feature extraction and ~~nonlinear~~non-linear feature extraction (Jain et al. 2000).

Linear feature extraction techniques include Principal Component Analysis (PCA), ~~Gray~~Grey level co-occurrence matrices (GLCM), Linear Discriminant Analysis (LDA), projection pursuit, and Independent Component Analysis (ICA). ~~Nonlinear~~Non-linear feature extraction methods include kernel PCA, PCA network, ~~nonlinear~~non-linear PCA,

~~nonlinear~~non-linear auto-associative network, Multi-Dimensional Scaling (MDS) and Self-~~Organizing~~Organising Map (SOM), Phase-Only Correlation (POC) and etc. (Zheng & He, 2005).

The best known linear feature extractor is the Principal Component Analysis (PCA) or Karhunen-Loève expansion that computes the *m* largest eigenvectors of the *d x d* covariance matrix of the *n d*-dimensional patterns. PCA uses the most expressive features, approximates the data by a linear subspace using the mean squared error criterion (Jain et al. 2000). The disadvantage of PCA is it does not take into account the vector's classes, so it does not look at the separation of classes ~~-~~ (Pascual, 2010).

~~Gray~~Grey Level Co-occurrence Matrix (GLCM) has been used extensively in the field of image processing. It has been applied from a range of applications like texture analysis to synthesis including ~~gray~~grey scale as well as ~~color~~colour texture recognition. The ~~Gray~~Grey level co-occurrence matrix (GLCM) computes the probability density function of image f (x,y) for all pair of pixels (i) and (j) in distance (d) with angular displacement (θ) = 0,45, 90 and 135 degree (Hawlick, 1979), the calculation will compute the frequency of ~~gray~~grey tone occurrence for angular adjacent pixels.

The GLCM is used for two purposes, quantizing the ultrasound images to improve the filtration, and extraction of features descriptors (Abdelrahman & Hamid, 2012). GLCM only contains co-occurrence information between two pixels, and thus cannot capture the spatial relationship between three or more pixels in the image. When all 256 grey levels are used in generating a GLCM, it shows every possible pixel pairs obtained.

This is a disadvantage as the GLCM consumes computational time to generate the matrices. Furthermore, the calculations of textural features from the GLCMs involve the calculation on every element of the GLCMs. ~~Hence~~As such, a lot of computational steps are wasted in these calculations. By selecting a lower grey level, it will reduce the size of the GLCM and computational time accordingly (Tou et al., 2008).

~~While Multidimensional~~Meanwhile the Multi-dimensional Scaling (MDS) is a ~~nonlinear~~non-linear feature extraction technique. It aims to represent a

~~multidimensional~~multi-dimensional data set in two or three dimensions ~~such~~so that the distance matrix in the original *d*-dimensional feature space is preserved in the projected space (Borg & Groenen, 2005). ~~A~~The problem with MDS is that it does not give an explicit mapping function, so it is not possible to place a new pattern in a map which has been computed for a given training set without repeating the mapping (Klock et al., 1998).

For the Self-Organizing Map (SOM), or Kohonen Map, neurons are arranged in an *m*-dimensional grid, where *m* is usually 1, 2, or 3. Each neuron is connected to all the *d* input units. During training, patterns are usually presented in a random order. One major problem with SOM is when getting the right data (Shinde & Deshmukh, 2011).

Unfortunately, a value for each dimension of each member of samples is needed in order to generate a map. This is simply not possible and often it is very difficult to acquire all of this data, so this is a limiting feature of SOM (Honkela, 1997) .

Another linear feature extraction is Independent Component Analysis (ICA). ICA has been successfully used for blind-source separation (Bingham, 2000). ICA is a statistical method for transforming an observed ~~multidimensional~~multi-dimensional random vector into components that are mutually as independent as possible.

The main limitation of ICA is that the extracted components should be regularly updated. This is because when someone is only looking for the presence of one component, all the extracted components have to be inspected (De Vos et al., 2006).

The phase-~~-~~only correlation technique also known as phase correlation function has been successfully applied to high-accuracy image registration task for computer vision application. The advantages of the phase-only correlation are their high discriminating power, numerical efficiency and robustness against noise (Zhang et al. 2006).

Whereas in biometric authentication tasks, POC shows an advantage since it is not influenced by image shift and brightness change (Ito, Aoki, Nakajima, et al., 2008). The most outstanding property of POC compared to the ordinary correlation is its accuracy in image matching. POC can also be extended to the registration of translated, rotated and scaled images (Takita et al., 2003).

There are three possible conditions that are possibly observed, namely: (a) if POC peak value of two images is "1", the first ~~images~~image is a copy of the second one, indicating that the images are from the same person, (b) if POC value of two images is lower than "1", but the POC peak can be observed, the two images may be considered as originated from the same object, (c) if POC peak of two images ~~is~~ does not appear, the two images can be considered as originated from different object (Arnia & Pramita, 2011).

A recognition system based on POC should ~~recognize~~recognise images of the same classes with higher correlation values, which is indicated by a present of POC peak. In opposition, it should ~~recognize~~recognise them of difference classes with lower correlation values without any POC peak (Arnia & Pramita, 2011).

Table 2.10 ~~summarizes~~summarises the ~~characterizations~~characterisations of each feature extraction technique as discussed. Some of the other feature extraction ~~technique~~techniques are also ~~state~~stated in this table.

Formatted: Indent: First line: 0 cm

Table 2.10: Feature Extraction Methods

Formatted: English (U.K.)

Formatted: English (U.K.)

Techniques	Property	Characteristic
Principal Component Analysis (PCA)	Linear map; fast; eigenvector-based	Traditional, eigenvector based method, also known as Karhunen-Loeve expansion, good for Gaussian data (Abdi & Williams, 2010)
~~Gray~~Grey level co-occurrence matrices (GLCM)	Homogeneity, ~~gray~~grey-tone linear dependencies, contrast,	~~also~~Also known as ~~Gray~~Grey level Dependency Matrix, identifying objects or regions of

	~~neighboring~~neighbouring co-occurrence matrix elements	interest in an image. (Gadkari, 2004).
Independent Component Analysis (ICA)	Linear map; iterative; non-Gaussian.	Blind source separation, used for de-mixing non-~~Guassian~~Gaussian distributed features (Comon, 1994)
Multidimensional scaling (MDS), and Sammon's projection	~~Nonlinear~~Non-linear map; iterative	Often poor ~~generalization~~generalisation; sample size limited; noise sensitive; mainly used for 2-dimensional ~~visualization~~visualisation (Webb, 1995).
Self-Organizing Map (SOM)	~~Nonlinear~~Non-linear ; iterative	Based on a grid of neurons in the feature space; suitable for extracting spaces of low dimensionality (Bullinaria, 2004) .
Phase-Only Correlation (POC)	~~Nonlinear~~Non-linear ; iterative	Invariant to illumination, rotation, and scaling (Ito et al., 2008),(Nakajima et al., 2008).

2.6.4 Feature Selection

The main purpose of this phase is to select the best subset from the input space. The aim is to select the optimal features subset that can achieve the highest accuracy results. Most of feature selection methods involve a ~~combinatorial search~~combination of searches through the whole space. Other methods divide the feature space into several ~~subspaces~~sub-spaces which can be searched easily (Zheng & He, 2005).

Feature selection used in many application areas as a tool to remove irrelevant or redundant features. The choice of the feature selection method depends on various data set characteristics~~. The characteristics include~~, including the data types, data size and noise.

In the data type's characteristic, there are two elements considered~~. The elements~~, which are features and class labels. The feature values can be continuous (C), discrete (D), or nominal (N). ~~While~~Meanwhile for the class label, some feature selection methods can handle only binary classes, and others can deal with multiple classes.

As for the data size, the feature selection deals with whether a method can perform well for small training set and whether it can handle large data size. Nowadays, ~~datasets~~data sets are mostly large in size so the second criterion is more practical and of interest to the current. The typical noises encountered during the feature selection process are misclassification and conflicting date~~.~~ (Dash & Liu 1997).

Basically, there are two types of feature selection methods~~. Which are~~ - filter and wrapper (Neumann et al. 2004). Filters act as a ~~preprocessing~~pre-processing step ~~independently~~independent of the classifier while wrappers take the classifier into account as a black box and embedded approaches that simultaneously determine features and classifier during the training process.

Table 2.11 lists most of the well-known feature selection methods. This table elaborates each method on their property and comments on the methods.

Formatted: Caption

Formatted: Font: Bold

Formatted: Font: Not Bold, English (U.K.)

Table 2.11: Feature Selection Methods

Techniques	Property	Characteristic
Exhaustive search	Evaluate all $\left(\frac{d}{m}\right)$ possible subsets.	Guaranteed to find the optimal subset; not feasible for even moderately large values of m and d.
Branch-and-Bound Search	Uses the well-known branch-and-bound search method; only a fraction of all possible features subsets need to be enumerated to find the optimal subset.	Guaranteed to find the optimal subset provided the criterion function satisfies the monotonicity property; the worst-case complexity of this algorithm is exponential.
Best Individual Features	Evaluate all the m features individually; select the best m individual features.	Computationally simple; not likely to lead to an optimal subset.
Sequential Forward Selection (SFS)	Select the best single feature and then add one feature at a time which in combination with the selected features ~~maximizes~~maximises the criterion function	Once a ~~features~~feature is retained, it cannot be discarded; computationally attractive since to select a subset of size 2, it examines only $(d - 1)$ possible subset.
Sequential Backward Selection (SBS)	Start with all the d features and successively delete one feature at a time.	Once a feature is deleted, it cannot be brought back into the optimal subset; requires more computation than sequential forward selection.
" *Plus l-take away r*" Selection	First enlarge the features subset by l features using forward selection and then delete r features using backward selection.	~~A voids~~Avoids the problem of features subset "nesting" encountered in SFS and SBS method; need to select values of l and r $(l > r)$.
Sequential Forward Floating Search (SFFS) and Sequential Backward Floating Search	A ~~generalization~~generalisation of "*Plus l-take away r*" method; the values of l and r	Provides close to optimal solution at an affordable computational cost.

Formatted: English (U.K.)

Formatted: Font: Italic

(SBFS)	are determined automatically and updated dynamically.	

Source: Jain et al. (2000)

The first two methods in the table guarantee an optimal subset while the others are ~~suboptimal~~sub-optimal caused by the fact that the best pair of features need not contain the best single feature. Generally, good, larger feature sets do not necessarily include good, small sets.

Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) operate by evaluating growing feature sets (forward selection) or by evaluating shrinking feature sets (backward selection). A simple sequential method like SFS and SBS, they add (SFS) and delete (SBS) one feature at a time.

The more sophisticated techniques are the "Plus l-take away *r*" Selection and Sequential Forward Floating Search (SFFS) and Sequential Backward Floating Search (SBFS). These methods backtrack as long as they find improvements compared to previous feature sets of the same size (Jain et al. 2000).

Formatted: Font: Italic

2.6.5 Classification

The ~~conceptualization~~conceptualisation of things, or objects, as belonging to classes is at the core of all knowledge. ~~Researcher~~The researcher's concern, however, is that classes are in the mind of the beholder and that there are often many possible attributes and ways to classify a set of objects. Commonly agreed upon classifications for many objects used routinely are a part of human cultures.

Pattern recognition involves two processes: (1) classification, where a sample from a population of objects is divided into groups called classes; and (2) recognition, where a given unknown object from the same population is ~~recognized~~recognised as belonging to one of the established classes. Recognition is sometimes divided into recognition and identification, which means that a particular individual object is ~~recognized.~~recognised. The terms classification and recognition are sometimes used interchangeably in the literature (Looney, 2003).

A classification process examines a sample of objects that represents a population of objects and partitions it into subsets (the classes) according to similarity of the objects within classes and dissimilarity of the objects between classes. This type of process is called self-~~organization~~organisation, unsupervised learning, or clustering of the sample into clusters (classes or ~~subclasses~~sub-classes).

~~On the other hand, once~~Once a process has clustered the sample into classes and each object is assigned a class label, an index value (or ~~codeword~~code word) that designates the particular class, a ~~recognizer~~recogniser can be trained to assign a class label to any unknown object from the same population (pattern recognition). The training process is called supervised learning or training of the ~~recognizer.~~recogniser. A trained ~~recognizer~~recogniser can perform pattern recognition online.

The classification stage is the decision making part of a recognition system and it uses the features extracted in the previous stage. The classifier component of a pattern recognition system has to be taught to identify certain feature vectors to belong to a certain class. This is because it is impossible to define the correct classes for all the possible feature vectors and the purpose of the system is to assign an object ~~which~~that it has not seen previously to a correct class.

It is important to distinguish between the two types of machine learning when considering the pattern recognition systems. The main learning types are supervised learning and unsupervised learning. We also use the terms supervised classification and unsupervised classification.

In supervised classification, we present examples of the correct classification (a feature vector along with its correct class) to teach a classifier. Based on these examples, that are sometimes termed prototypes or training samples, the classifier then learns how to assign an unseen feature vector to a correct class.

The generation of the prototypes has to be done manually in most cases. This can mean a lot of work. After all, it was because we wanted to avoid the hand-~~labeling~~labelling of objects that we decided to design a pattern recognition system in the first place. That is why the number of prototypes is usually very small compared to the number of possible inputs received by the pattern recognition system.

Based on these examples we would have to deduct the class of a never seen object. Therefore, the classifier design must be based on the assumptions made about the classification problem, in addition to prototypes used to teach the classifier. These assumptions can often be described best in the language of probability theory.

In unsupervised classification or clustering, there is no explicit teacher or training samples. The classification of the feature vectors must be based on similarity between them ~~based on~~, from which they are divided into natural groupings. Whether or not any two feature vectors are similar depends on the application. Obviously, unsupervised classification is a more difficult problem than supervised classification and supervised classification is the preferable option if it is possible (Medlock & Briscoe, 2007), (Barbu, 2006).

In some cases, however, it is necessary to resort to unsupervised learning. For example, this is the case if the feature vector describing an object can be expected to change with time. There exists a third type of learning: reinforcement learning. In this learning type, the teacher does not provide the correct classes for feature vectors but the teacher provides feedback on whether the classification was correct or not (Tohka, 2011).

Designing a classifier is based on feature and models selected (Shinde & Deshmukh, 2011). If ~~classifier~~the classifier's performance is not good, the model has to be changed. The classifier is evaluated with the testing data and error rate is determined.

~~Criteria~~The criterion for accepting or rejecting a classifier ~~are~~is based on the performance of the classifier for the given set features. The feature vector obtained from the previous step is passed to a classifier that evaluates the evidence presented and assigns the object to a class (Verma & Raja, 2005).

After the optimal feature subset is selected, a classifier can be designed using various approaches. There are three different approaches involved. The first approach is the simplest and the most intuitive~~ approach~~, which is based on the concept of similarity. The examples include Template matching and 1-Nearest ~~Neighbor~~Neighbour Rule (1-NN~~) are the examples~~).

The most straightforward 1-Nearest ~~Neighbor~~Neighbour Rule can be used as a benchmark for all the other ~~classifier~~classifiers, as it appears to always provide a reasonable classification performance in most applications. 1-NN classifier does not require any user-specified parameters and its classification results are implementation independent (Jain et al., 2000). As more invariants are considered, the dimensionality of ~~subspace correspondingly~~sub-space increases in tandem. Template matching the nearest mean, classifier can be viewed as finding the nearest ~~subspace.~~sub-space.

The second approach is a probabilistic approach. It includes methods based on Bayes decision rule, the maximum likelihood or density estimator. Two well-known are K-nearest neighbour (KNN) and Parzen window classifier (Jain et al. 2000)~~(Zheng & He 2007)~~ (Zheng & He, 2005).

KNN is a typical instance-based prediction model~~. By KNN,~~, where the class label of a new testing sample is decided by the majority class of its k closest neighbour based on their Euclidean distance (Dash & Liu, 1997). KNN classification is one type of lazy classification algorithm that offers many advantages.

The ~~k~~K-nearest neighbour (~~k-NN~~KNN) technique is a classic, simple and appealing method to address classification problems. It has a very intuitive interpretation and its predictions are easily explained to domain experts. As a result, KNN has been applied for classification in many domains.

~~K-NN~~KNN is a type of instance-based learning where the function is only approximated locally and all ~~computation is~~computations are deferred until classification. In ~~K-NN~~KNN, an object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbour. The neighbours are taken from a set of objects for which the correct classification is known. This can be thought of as the training set for the algorithm, though no explicit training step is required (Gnanasivam & Muttan, 2012).

The ~~neighbors~~neighbours are taken from a set of samples for which the correct classification is known. It is ~~usual~~common to use the Euclidean distance, though other distance measures, such as the City block, Cosine distances could be used instead (Guru et al., 2010). In pattern recognition, the ~~k~~K-nearest neighbour algorithm (~~K-NN~~KNN) is the generally used method for classifying objects based on closest training examples in the feature space.

The Parzen Window classifier is replacing the class~~—~~-conditional densities. Both of these classifiers require the computation of the distances between a test pattern and all the patterns in the training set (Jain et al., 2000).

The third approach is to construct decision boundaries directly by ~~optimizing~~optimising certain error criterion. The examples are Fisher's Linear Discriminant, Multi-layer Perceptron (Feed-Forward Neural Network), and Binary Decision Tree (Zheng & He, 2005).

Fisher's Linear Discriminant ~~minimizes~~minimises the mean squared error (MSE) between the classifier output and the desired labels. ~~While~~Meanwhile the single-~~–~~layer perceptron, where the separating hyperplane is iteratively updated as a function of the distances of the misclassified patterns from the hyperplane, if the sigmoid function is used in combination with the MSE criterion, as in feed-forward neural nets (also called Multi-layer Perceptron), the perceptron may show a ~~behavior which~~behaviour that is similar to other linear classifiers.

A Naive Bayes classifier is a simple probabilistic classifier based on Bayes' theorem from Bayesian statistics with strong independence assumptions. A more descriptive term for the underlying probability model would be an independent feature model. Thus, the Naive Bayes independence model is based on estimation (Nilsson, 1965).

In simple terms, a naive Bayes classifier assumes that the presence or absence of a particular feature of a class is unrelated to the presence or absence of any other feature. Depending on the particular nature of the probability model, the naive Bayes classifiers can be trained very competently in a supervised learning setting.

The advantage of the naive Bayes classifier is that it has a short computational time for training. In addition, since the model has the form of a product, it can be converted into a sum through the use of logarithms with significant consequent computational advantages.

If a feature is numerical, the usual procedure is to separate it during data pre-processing (Yang & Webb, 2001), although a researcher can use the normal distribution to calculate probabilities (Bouckaert, 2004).

Other than that, another special type of classifier is the decision tree, which is trained by an iterative selection of individual features that are most salient at each node of the tree. During classification, just those features are under considerations that are needed for the test pattern under consideration (Jain et al. 2000). The details of those classifiers are ~~summarized~~summarised in Table 2.12.

Table 2.12: Classification Methods

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Techniques	Property	Characteristic
Nearest Mean Classifier	Assign patterns to the nearest class mean.	Almost no training needed; fast testing; scale (metric) dependent- (Looney, 2003) .
Subspace Method	Assign patterns to the nearest class subspace.	Instead of normalizing on invariants, the subspace of the invariants is used; scale (metric) dependent (Ho, 1998).
1-Nearest ~~Neighbor~~Neighbour Rule	Assign patterns to the class of the nearest training pattern.	No training needed; robust performance; slow testing; scale (metric) dependent (Jain et al. 2000).
k-Nearest ~~Neighbor~~Neighbour Rule	Assign pattern to the majority class among k nearest ~~neighbor~~neighbour using a performance optimized value for k.	Asymptotically optimal; scale (metric) dependent; slow testing (Hawlick 1979) .

Bayes plug-in	Assign pattern to the class which has the maximum estimated posterior probability.	Yields simple classifiers (linear or quadratic) for Gaussian distributions; sensitive to density estimation errors (Jain et al. 2000).
Parzen Classifier	Bayes plug-in rule for Parzen density estimates with performance optimized kernel.	Asymptotically optimal; scale (metric) dependent; slow testing (Jain et al. 2000).
Fisher Linear Discriminant	Linear classifier using MSE optimization.	Asymptotically optimal; scale (metric) dependent; slow testing (Zheng & He 2005).
Binary Decision Tree	Finds a set of thresholds for a pattern-dependent sequence of features.	Iterative training procedure; overtraining sensitive; needs pruning; fast testing (Zheng & He, 2005).
Perceptron	Iterative optimization of a linear classifier.	Sensitive to training parameters; may produce confidence values (Zheng & He 2005).
Multi-layer Perceptron (Feed-Forward Neural Network)	Iterative MSE optimization of two or more layers of perceptrons (neurons) using sigmoid transfer functions.	Sensitive to training parameters; slow training; nonlinear classification function; may produce confidence values; overtraining sensitive; needs regularization (Silva et al. 2011).
Support Vector Classifier	Maximized the margin between the classes by selecting a minimum number of support vectors.	Scale (metric) dependent; iterative; slow training; nonlinear; overtraining insensitive; good generalization performance (Michie et al. 1994).
Naive Bayes Classifiers	Compare two probabilities by using a product operation	Short computational time for training, converted into a sum through logarithms (Kotsiantis, 2007)

2.6.6 Post – Processing

The final task of the pattern recognition system is to decide upon an action based on the classification result (Tohka, 2011). In most pattern recognition systems, ~~some~~ data processing is performed after the classification stage. These post processing, such as the normalization processes, bring some prior information about the surrounding world into the system.

These additional processes can be ~~utilized~~utilised in improving the overall classification accuracy. The processing phase is possible if the individual objects or segments form ~~as~~ meaningful entities such as bank account number, words or sentences. The soundness or existence of these higher level objects can be examined and if an error is indicated, further step can be taken to correct the misclassification. The post processing phase thus resolves interdependencies between individual classifications (Lampinen, 1997).

2.7 SUMMARY

Pattern Recognition plays a ~~very~~ vital role in Artificial ~~intelligence. But recently~~Intelligence. Recently pattern recognition has become a routine in ~~everyday's~~daily life. As ~~human beings~~humans have limitations in ~~recognizing~~recognising various items, the field of pattern recognition is becoming very popular. The goal of pattern recognition research is to clarify complicated decision making processes and ~~automated~~automate these ~~function~~functions using computers ~~for~~in our daily life.

Pattern recognition has various applications in numerous fields ~~as~~, including data mining, biometrics, sensors, speech recognition, medical, military, natural language processing etc. Statistical pattern recognition is used to cover all stages of an investigation, from problem formulation and data collection through discrimination and classification, to assessment of results and interpretation. Here each pattern is represented in terms of features and is viewed as a point in a d-dimensional space.

Currently, the timber industry is one of the major contributors to the national economy. To ensure that the industry ~~can~~could continue to ~~increase~~grow, the quality of wood used should be high and appropriate for its usage.

In this highly demanding timber industry, wood species recognition is widely used in various areas such as in construction industry and manufacturing industry. Each species of wood has its own characteristics and features, which fit the use in certain industries. For example, ~~for construction,~~ hardwood such as Balau, Chengal and other are preferably used~~.~~ for the construction sector. The misuse of wood might affect the result or products and ~~can allow danger to the consumer.~~could endanger consumers.

Currently, recognition of wood species in Malaysia is done manually. There are two ways of ~~recognizing the~~ recognising wood species ~~which are by~~ using the naked ~~eyes~~eye and with the aid of magnifier. ~~When the~~The recognition process of the naked eyes~~, the recognition process~~ is usually done without any aid of any accessories and ~~it~~ is manually done by only the experts.

~~While when recognizing~~Meanwhile, for the ~~woods by~~recognition with the aid of a magnifier, it is compulsory to have a folding type 10x magnification magnifier. Besides that, the right cross section technique is required to make a good clean surface by using a sharp pocket knife. This is important to obtain well defined features of wood.

This research focused on Malaysian wood species recognition ~~which purposed~~aims to ~~computerize~~computerise the process of Malaysian wood species recognition. Several studies have been done to identify the suitable techniques to solve the research problems.

There are several techniques that may be used for pattern recognition such as Principle Component Analysis (PCA), Independent Component Analysis (ICA), Multidimensional Scaling (MDS), Phase-Only Correlation (POC) and others. From the studied techniques, POC has been ~~chosen to be performed in~~selected for the feature extraction phase ~~for~~of wood species recognition.

The selection is based on the advantages of POC as it is highly ~~power~~effective for discrimination, numerical efficiency, ~~robustness~~robust against noise and not influenced by image shift and brightness change as compared to other techniques ~~state~~stated in the literature review.

Meanwhile, for the classification technique, the KNN ~~technique is~~method has been chosen as it is a statistical classifier. ~~They~~It can predict class membership probabilities, such as the probability that a given sample belongs to a particular class.

From the studies studied, when comparing the classification algorithms it is found that the KNN technique is more suitable in performance compare ~~with~~to decision tree and selected neural network classifiers. ~~Besides that~~Furthermore, KNN have also exhibited high accuracy and speed when applied to large databases.

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

This chapter describes the methodology used in this research. In this chapter, four (4) phases have been used to develop wood species recognition system. The framework describes the detail of every phase in developing the wood species recognition system.

3.2 FRAMEWORK OF THE SYSTEM

In this study, the development of a prototype for wood species recognition system is divided into four main phases ~~such as~~including image acquisition, pre-processing, feature extraction and features analysis. ~~Whereby, each~~Each of these phases will have ~~their~~its own module. Figure 3.1 ~~illustrated~~illustrates the framework for the wood species recognition system.

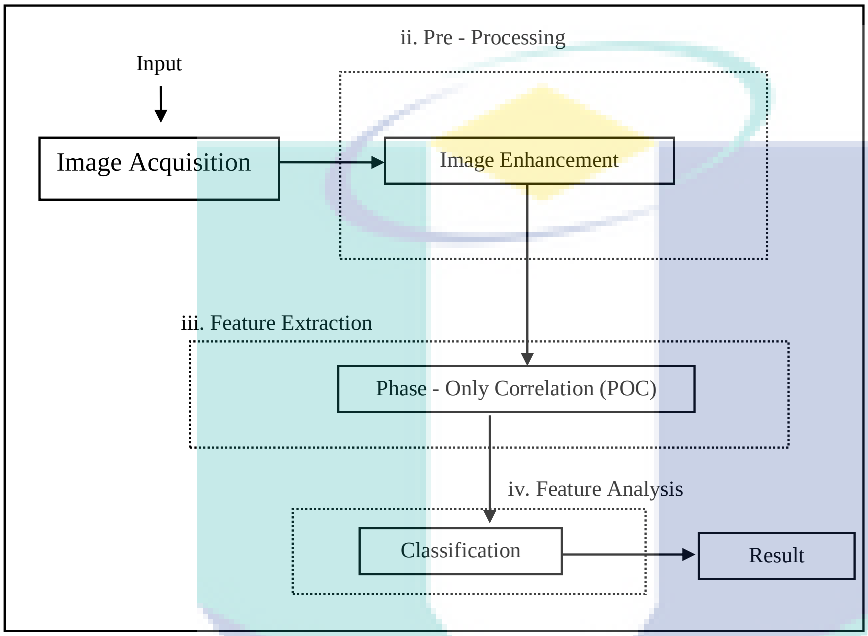


Figure 3.1: Framework of wood species recognition system

Every phase in this system is divided into several small processes that perform an important role to ensure that application development's goals are achieved. The first process of this system is image acquisition phase where the input samples are captured using a camera. Generally, this phase is served to obtain images ~~of the~~for applications purposes.

~~Pre~~The pre-processing process will be applied after image acquisition process is ~~finished.~~completed. After the input image is captured, it will be ~~normalized~~normalised through the pre-processing process. The purpose of pre-processing is to improve the quality of the image processed. It makes the subsequent phase of image processing like recognition of wood species easier.

Formatted: hps

Formatted: Font:

Formatted: Font:

Formatted: Font:

Formatted: Font:

Whereas, feature extraction phase is intended for extracting the wood species image features. The ~~normalized~~normalised input will be used for feature extraction where the important textural features are extracted here. The extracted features will be classified using ~~a~~the phase analysis and produces an output as shown in Figure 3.1.

Phase analysis is the final phase in this system~~. This phase is the phase of the determinant~~, which is to determine whose role is to ~~recognize~~ recognise the types of woods species. Figure 3.1 shows a sketch of the whole process involved in the wood species recognition system.

3.3 IMAGE ACQUISITION

Image acquisition is the first step in image processing. It is intended for wood images to be used as unprocessed data for the wood species recognition system. In addition, this process aims to convert images from digital ~~image~~ to ~~computerize~~computerised image.

For the wood species recognition system, wood images are taken by the author using a Canon IXY 200F digital camera. There are 1,280 images ~~are~~ taken with each size of ~~it has a~~ 450 x 300 pixels and they are saved as JPEG files. The JPEG file format is used in this system because it is appropriate to use compared to other formats. JPEG algorithm can provide a good image quality for the image file size and ~~bit~~can be deficient if a small image is used (Umbaugh 1998).

Other than that, ~~grayscale~~grey scale image ~~are~~ is also used in this study~~. Grayscale image representation used by the appropriateness of the~~ as it provides appropriate wood image ~~used~~for use as unprocessed data for the wood species recognition system. ~~Thus, grayscale~~Grey scale image has only one ~~color~~ colour or monochrome. Therefore, it is easily processed and ~~analyzed~~analysed rather than ~~image colour~~ coloured images~~ which~~, as

Formatted: Space Before: 0 pt, After: 10 pt they require more work to ~~analyze~~analyse the three basic ~~colors that there are~~colours - red, blue and green.

For the development of the whole system, the input images are captured using the camera. ~~Wood~~The wood image data must be identified and ~~analyzed prior~~analysed before it was tested. Knowledge of the form features for each type of wood species should be obtained~~. So~~ so that~~,~~ the wood image species can be determined. This is important to determine whether each of the data will get the results they deserved. In this study, eight (8) wood species are used as the unprocessed data. Figure 3.2 shows samples of wood species.

Formatted: hps

Wood samples of eight (8) wood species were collected from the Malaysian Timber Industry Board in Indera Mahkota Pahang, namely Balau, Mersawa, Ramin, Mersawa Gajah, Durian, Rengas, Melunak and Sepetir. Basically, wood species can ~~recognize~~be recognised based on two kinds of features, which are physical ~~features~~ and anatomical ~~features~~.

~~In~~For physical features, wood species are ~~recognized~~recognised based on the colour, texture and appearance of specimens, and touch was used to determine whether the wood grains are rough, smooth, dull, lustrous or glossy, whether the wood is light or heavy and also the kind of smell produced by the wood species.

Meanwhile, anatomical feature ~~used~~uses three key elements of wood structure to ~~recognize~~recognise the wood species. The three key elements are rays, vessels and parenchyma. Figure 3.3 to Figure 3.10 ~~illustrates sample of wood~~ illustrate the anatomical features~~.~~ of the wood samples.

Formatted: hps

Formatted: Left

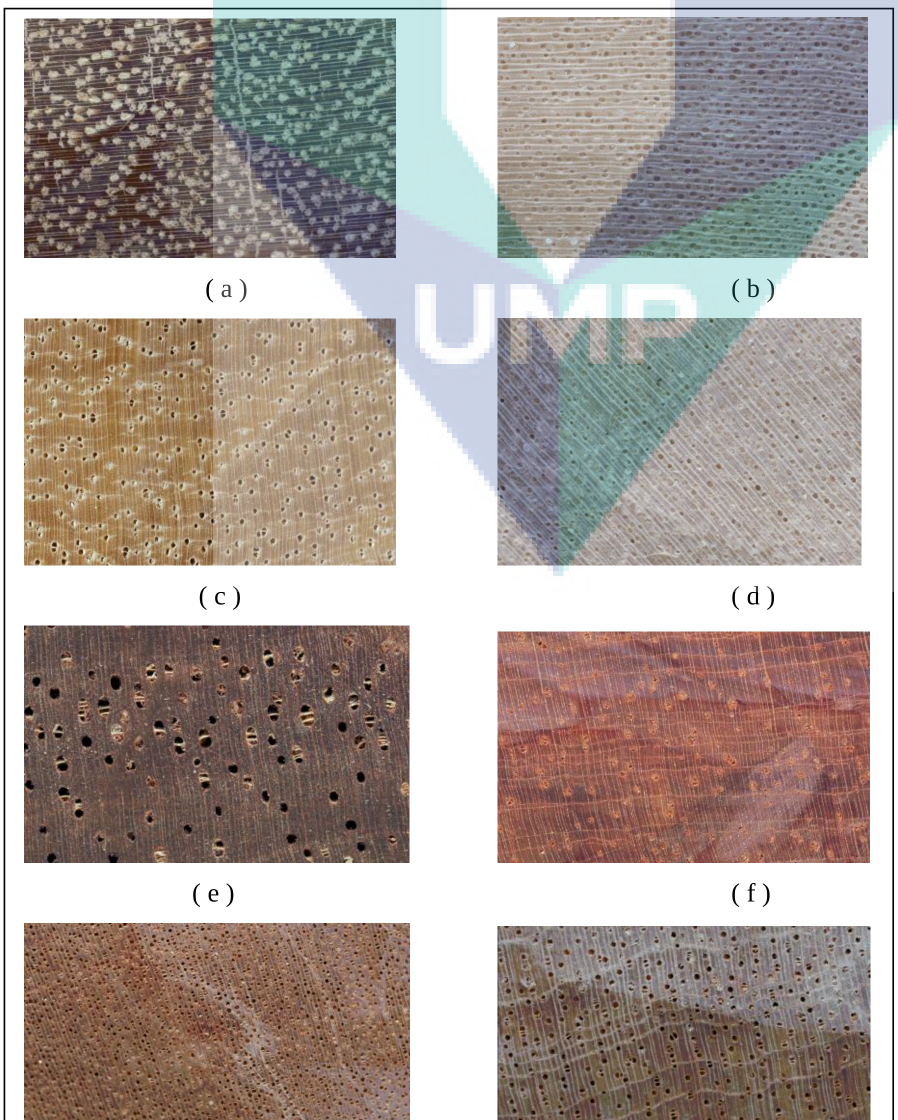
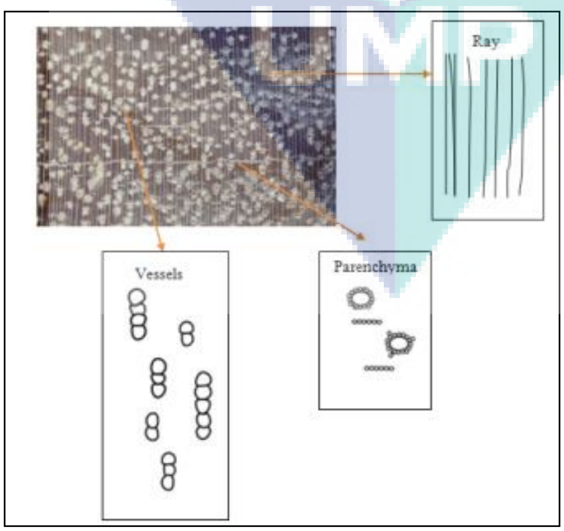


Figure 3.2: Eight Wood Species (a) Balau, (b) Mersawa, (c) Ramin, (d) Mersawa Gajah, (e) Durian, (f) Rengas, (g) Melunak and (h) Sepetir.

Each species has different type of vessels, rays and parenchyma. The elements ~~are~~ differed according to the species. Figure 3.3 shows an image of Balau species and its rays, parenchyma and vessels. Balau appears with simple perforation and mostly solitary vessels while ~~it~~its parenchyma is usually diffuse strand and mostly incomplete vasicentric. The rays for Balau are fine and visible to the naked eyes.



Formatted: English (U.K.)

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Figure 3.3: Wood elements for Balau

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

~~While~~Meanwhile Figure 3.4 shows an image of the Ramin species~~. Ramin~~, which consists of generally medium-sized and solitary vessels. The parenchyma is exclusively aliform with extended wings and its rays are fine but not visible to the naked eye. In Figure 3.5, Durian shows an interesting vessel when the vessels are mostly large and ~~it~~ is solitary and has multiple pores. Its parenchyma is diffuse strands or occasionally in closely spaced. The rays of Durian species consist of two distinct sizes, of which ~~are~~ the finer rays are not visible to the naked eye while the broader medium-sized rays are distinct.

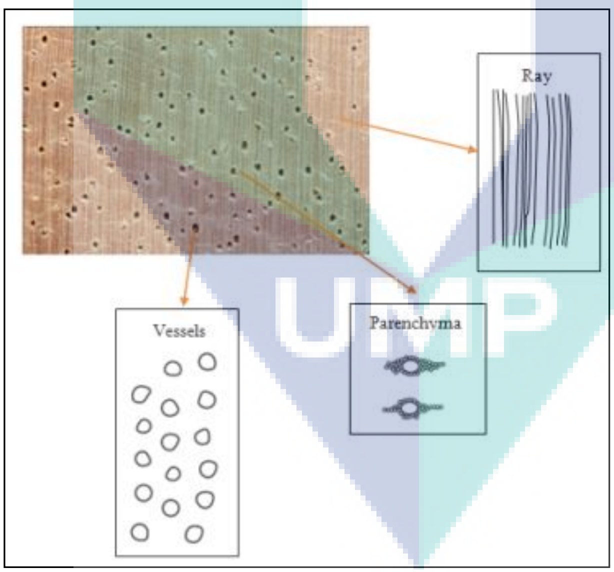


Figure 3.4: Wood elements for Ramin

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Figure 3.5 shows an image of the Melunak species and its vessels, parenchyma and rays pattern. For the vessels, Melunak appears with medium-sized, solitary and occasionally ~~clusters~~clustered vessels and the parenchyma is occasionally aliform or locally confluent. The rays of Melunak species are moderately fined, just visible for the naked eye on cross section.

Figure 3.6 shows an image of the Mersawa wood species. ~~For Mersawa, the~~Its vessels are medium-sized or moderately large with solitary vessels~~. Its~~ while its parenchyma is vasicentric in pattern. The rays ~~are~~ exist in two sizes; the broader rays medium-sized are distinct to the naked eye in cross section while the interspersed very fine rays are only visible ~~only~~ with a hand lens.

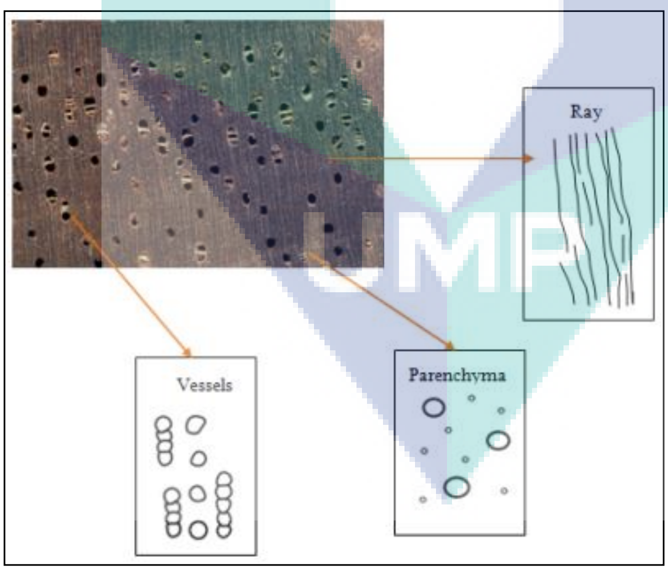


Figure 3.5: Wood elements for Durian

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

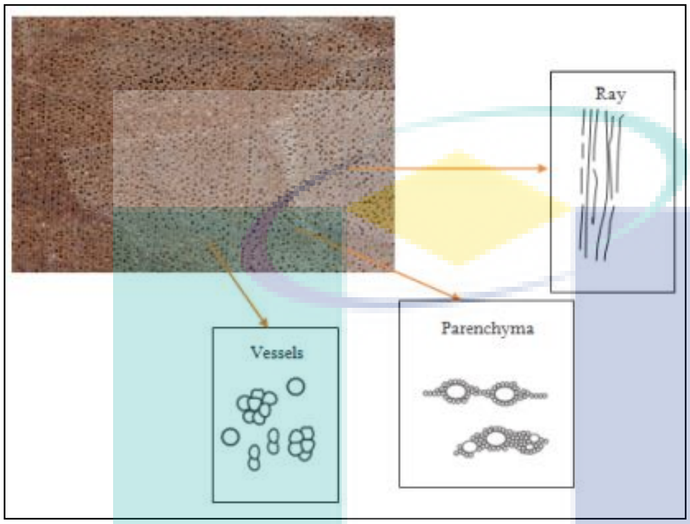


Figure 3.6 : Wood elements for Melunak

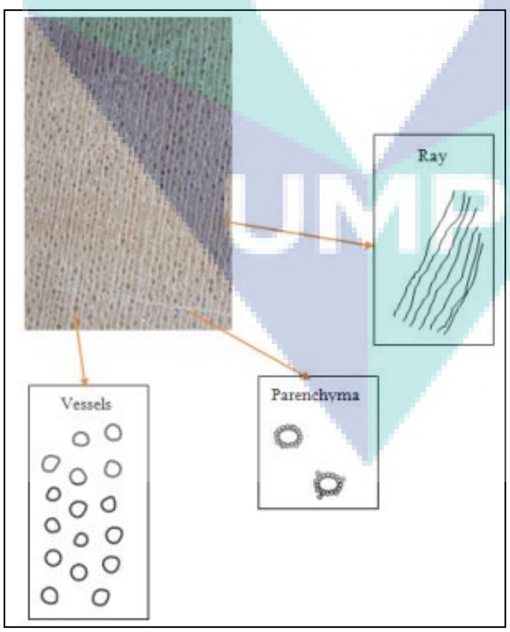


Figure 3.7: Wood elements for Mersawa

Formatted: Font: Bold, English (U.K.)
Formatted: English (U.K.)

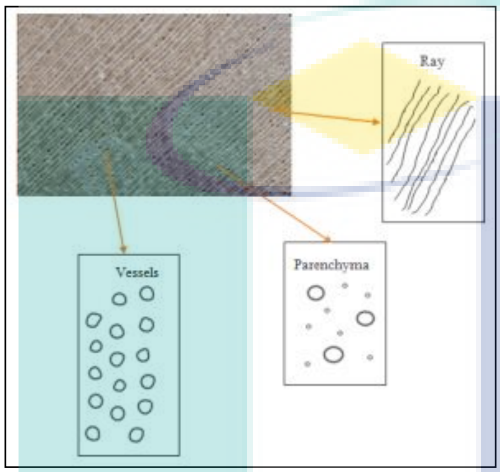


Figure 3.8: Wood elements for Mersawa Gajah

Formatted: Font: Bold, English (U.K.)
Formatted: English (U.K.)

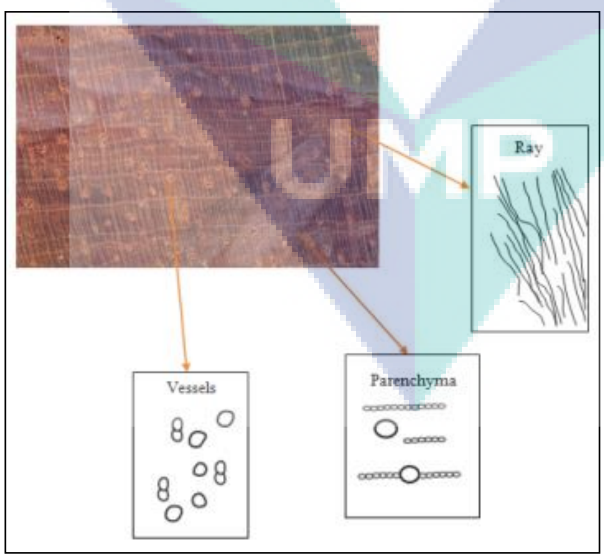


Figure 3.9: Wood elements for Rengas

Formatted: Font: Bold, English (U.K.)
Formatted: English (U.K.)

Different ~~with~~from Mersawa gajah in Figure 3.7, the vessels are medium-sized or moderately large with solitary vessels but the parenchyma is mostly in diffuse strands. The rays~~ are~~ also exist in two sizes; the broader rays medium-sized are distinct to the naked eye in cross section while the interspersed very fine rays are only visible~~ only~~ with a hand lens.

Figure 3.8 shows an image of the Rengas species. The vessels of Rengas species are moderately large, have solitary pores and diffused strands without any arrangement~~. While~~, while the parenchyma is banded. The rays are fine but generally not visible to the naked eye. In Figure 3.9, an image of the Sepetir species shows that the vessels of Sepetir are mostly moderately large in size, have solitary pores and diffused strands in pattern. The parenchyma is vasicentric and ~~tending to~~tends towards the aliform type. The rays are moderately fine and generally visible to the naked eye.

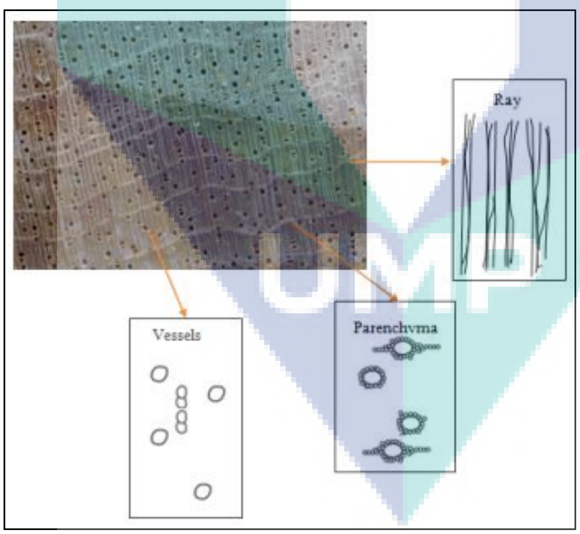


Figure 3.10: Wood elements for Sepetir

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

3.4 PRE - PROCESSING IMAGE

The second phase of the wood species recognition system is image enhancement, which is known as the pre~~ - ~~-processing phase. Image enhancement improves the quality of images for human perception by removing noise, reducing ~~blurring~~blurriness, increasing contras and providing more detail. This process is important to produce a more effective input to the feature extraction phase. Figure 3.11 ~~illustrate~~illustrates the processes that occur in pre~~ - ~~-processing image phase for wood species recognition system.

Firstly~~ ~~, the wood images are taken by using a digital camera~~. Therefore, ~~, which often caused some of the side effects ~~often ~~to occur. Sometimes, uneven ~~light effect~~lighting occurred in the image ~~caused the image seems~~causing it to be partly dark and partly clear. This situation will ~~cause~~lead to the density of ~~the ~~different ~~colors~~colours of pixels in the resulting image. As a solution ~~of~~to the problem, image enhancement is used to remove noise or correct the contrast in the image (Alginahi, 2010).

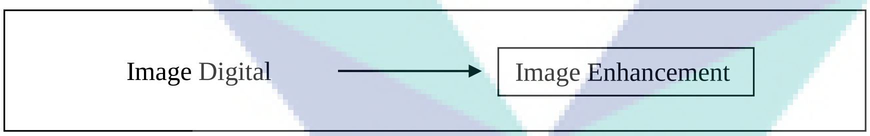


Figure 3.11: Pre - Processing for Wood Species Recognition System

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

In order to enhance the image, average filtering has been used to remove the noise of the image. The ~~details~~detailed examination ~~about~~on these techniques will be discussed in the following sections.

3.4.1 Average Filtering

Generally, the basic approach used for filtering technique is the sum of window coefficients and the density of pixel values inside the window at an image location method (Dzulkifli, 1997).

~~Assume the~~ The size assumed of the common window is 3 x 3 images as shown in Figure 3.12 (c). The values of the ~~gray~~grey levels are expressed as P_1, P_2 ... P_9 and coefficients window specified as M_1, M_2 ... M_9. By the action of the window linear spatial, S is:

$$S = P_1M_1 + P_2M_2 + \ldots + P_9M_9 \tag{3.1}$$

~~Generally~~The average filtering technique ~~is considering~~generally considers the average derived from N x N window (Figure 3.12 (b)). It uses the concept of window's changing to process the entire image. This is the average filtering arithmetic obtained from this Equation (Umbaugh, 1998):

$$s = \frac{1}{N^2}\sum_{(r,c)\in W} d(r,c) \tag{3.2}$$

~~with~~With N^2 is the number of pixels in the window N x N of W. The first step collects the original image ~~gray~~grey level (Figure 3.12 (a)) after image acquisition done. ~~Gray~~Grey level pixel obtains through original ~~gray~~grey level image (Figure 3.12 (b)).

Formatted: hps

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

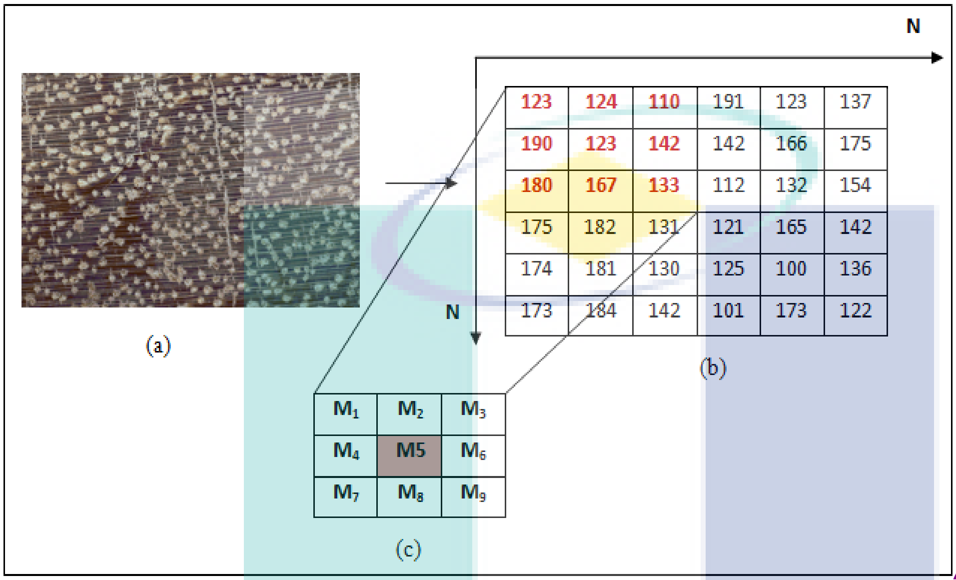


Figure 3.12~~:~~: ~~Illustrate~~Illustrates the image enhancement process for Average filtering technique~~.~~: (a) ~~gray~~grey level image, (b) ~~gray~~grey level pixel, (c) windows 3 x 3 pixel

~~While for~~ Meanwhile, the 3 x 3 window is scanned from left to right and top to bottom in overall image. Finally, each value of the pixel in the window 3 x 3 is converted ~~to~~by the calculation of a new value, *S* as Figure 3.12 (c). The example of calculation is~~-~~:

$$S = M_1 + M_2 + M_3 + M_4 + M_5 + M_6 + M_7 + M_8 + M_9 /9$$

$$= 123 + 123 + 124 + 110 + 190 + 142 + 180 + 167 + 133 / 9$$

$$= 143.556$$

The calculation above is repeated for every pixel in the original image to generate the smoothed image.

Formatted: English (U.K.)

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

3.5 FEATURE EXTRACTION

The next phase for wood species recognition system is feature extraction. Generally, the purpose of image extraction phase in this study is to extract the features that are available in the wood image. ~~Extraction~~The extraction process will produce features of different images for different species.

This process reduces the input values by ignoring the raw pixel values of the whole input images and ~~focus~~focuses only on useful values that ~~represents~~represent the features or properties of the image. In this research, phase~~-~~only correlation (POC) ~~were~~was used to extract the feature of every wood species.

3.5.1 Phase~~-~~-Only Correlation (POC)

Phase~~-~~Only Correlation ~~principal,~~(POC) principally has three steps that are directly involved in the process, namely 2D Discrete Fourier Transform (2D DFT), cross~~-~~-phase spectrum or ~~normalized~~normalised spectrum and 2D Inverse Discrete Fourier Transform (2D IDFT). All of the techniques are important process in the phase~~- only~~-sonly correlation function. Figure 3.13 ~~illustrate~~illustrates the processes that occur in phase~~-~~-only correlation method for the wood species recognition system.

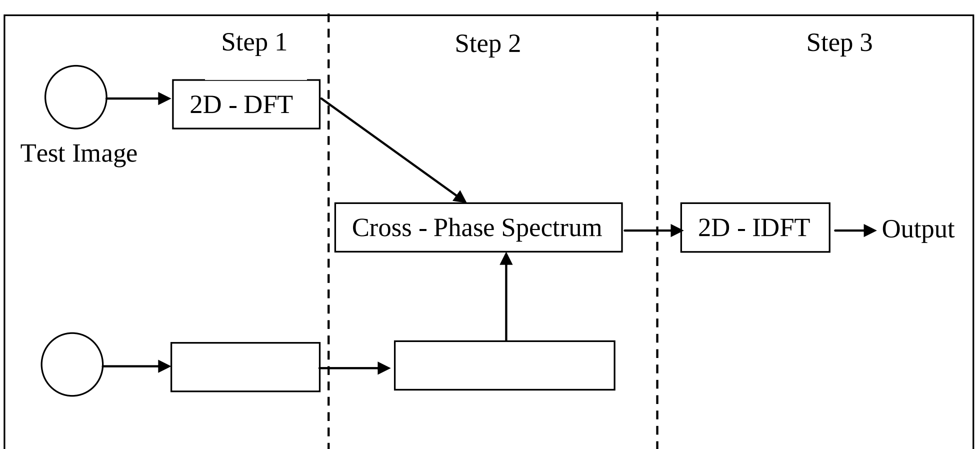


Figure 3.13: Diagram for Implementation of Phase - Only Correlation (POC)

Source : Rao et al. (2010)

Formatted: English (U.K.)

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Step 1: 2D Discrete Fourier Transforms (2D DFT)

Step 1 used 2D Discrete Fourier Transforms (2D DFT) .The purpose of 2D Discrete Fourier Transforms (2D DFT) is to calculate the frequency of images. Consider two (2) images N_1 x N_2 which are training image as $f\ (n_1,n_2)$ and test image as $g\ (n_1,n_2)$. Meanwhile, ~~for~~the 2D Discrete Fourier Transforms (2D DFT) ~~declare~~is declared as $F\ (k_1,\ k_2)$ and $G\ (k_1,\ k_2)$.

For training, ~~grayscale~~grey scale images with 256 bins are used. Firstly, frequency of every bin is computed and stored in vectors for further processing. Secondly, the mean of the consecutive nine (9) frequencies as $f\ (n_1,\ n_{2,}\ _{\dots}\ ,\ n_9\ )$ from the stored vectors is ~~calculate~~calculated and stored in another vectors for later use in the testing phase. Equation 3.3 define of total frequencies of the nine (9) consecutive bin in the histogram to the number of bin used to output Y. Figure 3.14 ~~illustrate~~illustrates the training algorithm and Figure 3.15 ~~shown~~shows the training process histogram.

$$Y = \frac{number\ frequencies}{total\ number\ of\ frequencies} \tag{3.3}$$

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Procedure: Training

Input: Eighty (80) images from every eight (8) subject of wood database

Start Training

1. Calculate frequency of every bin and store in matrix.

2. Calculate mean of consecutive nine frequencies and store in matrix for next step of comparison.

End Training

Figure 3.14 : Training algorithm

Formatted: English (U.K.)

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

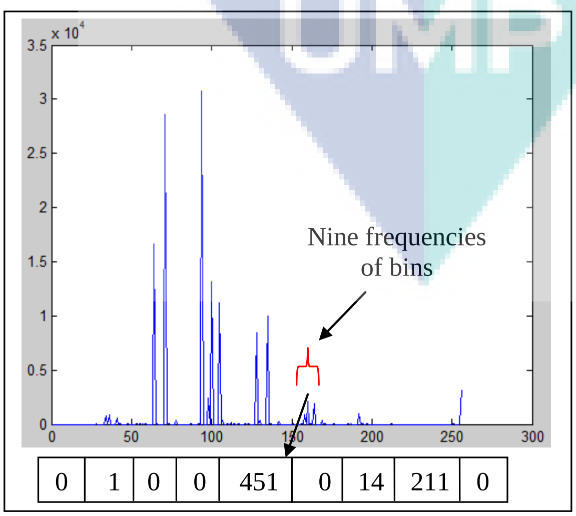


Figure 3.15: Training process histogram

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

The peak of each bin shows the frequency of that particular bin. To get the training data for this system, the mean of nine (9) frequencies of bins for each wood species is computed. Once the data is computed, it is stored in the database as ~~the~~ training data. The example mean of the consecutive nine frequencies calculation is shown below:

Formatted: Indent: First line: 1.27 cm

$$Y = \quad 0 + 1 + 0 + 0 + 451 + 0 + 14 + 211 + 0 / 9$$
$$= \quad 75.222$$

Formatted: Font: Italic

Figure 3.16 illustrates the chunk of bins found from training data. Figure 3.16-(a) shows the frequencies of 256 bins used ~~and the~~while Figure 3.16(b) shows the mean of nine frequencies of training images computed. This mean vector is used for calculating the absolute differences among the mean of trained images and test image. Finally the minimum ~~difference~~differences found ~~identifies~~identified the matched class with test image.

Similarly, the preliminary steps of testing are the same as training. Furthermore the POC values between test image and images of predicted class are computed for match or mismatch. The propose ~~algorithm~~algorithms used for testing are shown in Figure 3.17 respectively.

0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
11	36	62	220
0	0	0	0
2306	317	2411	1645
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
6208	12091	7324	14866
0	0	0	0
332	13	277	87
0	0	0	0
0	0	0	0

257.4444	39.22222	274.7778	207.2222
726.6667	1344.889	844.5556	1661.444
13.55556	62.66667	22.44444	36.11111
8.333333	0.777778	1.666667	0.222222
4618.333	4230.556	4358.111	4358.556
41.77778	44.44444	48	36.88889
0	0.444444	0.222222	0.222222

(b)

(a)

Figure 3.16: Sample of training data

Procedure: Testing

Input: Eighty (80) images from every eight (8) subject images of UMP
(Not used in Training)

Output

1. Absolute processed histogram differences.
2. POC value with match subject.

Begin Testing

1. Calculate frequency of every bin of test image.

2. Calculate mean of consecutive nine frequencies of test image.

3. Compare the means of test image with all trained matrix mean values in two (2) of training algorithm.

4. Identify the image with minimum absolute difference.

5. Calculate POC value with that identified image in step four (4).

6. Compare POC values with training image and test image for verification.

End Testing

Figure 3.17: Testing Algorithm

Figure 3.18 illustrates the sample of testing data. Figure 3.18(a) shows the frequencies of 256 bins while Figure 3.18(b) shows the mean of nine (9) frequencies~~.~~. The testing data underwent the same process as the training data. The set of computed mean values are used for calculating absolute differences between the means of trained images and testing images.

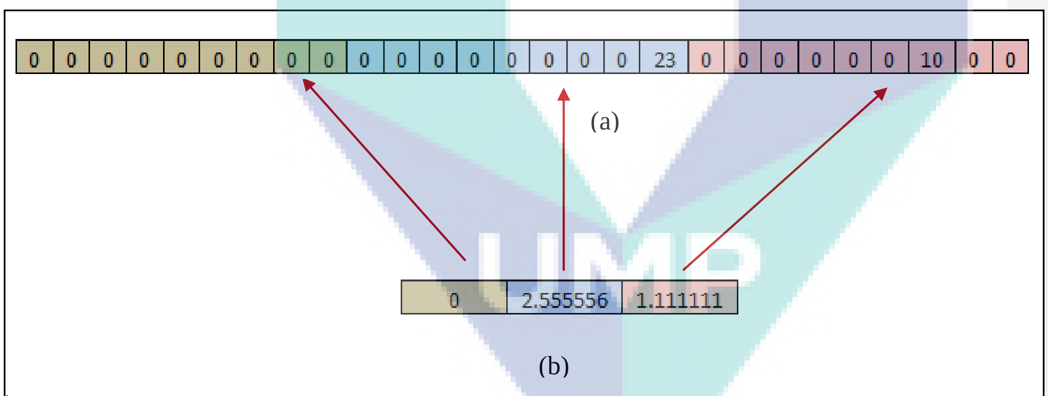


Figure 3.18: Sample of testing data

Step 2: Cross - Phase Spectrum

In step 2, two (2) images are used namely training image and test image, with the cross-phase spectrum technique ~~measures~~measuring the similarity ~~for~~of each translation in

Formatted: English (U.K.)

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

an image patch. The correlation must be ~~normalized~~normalised to avoid contributions from local image.

The cross-phase spectrum $R_{FG}(k_1, k_2)$ (Eq. (3.4)) implies that the calculation of POC emphasizes on the high frequency component. ~~Then~~The cross~~-~~-phase spectrum R_{FG} is ~~define~~then defined as the conjugate of $G\ (k_1,\ k_2)$ multiplied by $F\ (k_1,\ k_2)$ divided by its absolute value ~~as follows.~~(refer to the following).

$$R_{FG}(k_1, k_2) = \frac{F(k_1, k_2)\,\overline{G\ (k_1, k_2)}}{\left|F(k_1, k_2)\,\overline{G\ (k_1, k_2)}\right|}$$

$$= e^{j\theta(k_1, k_2)} \qquad (3.4)$$

Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)

Step 3: 2D Inverse Fourier Transform (2D IFT)

The 2D Inverse Fourier Transform (2D IFT) is to eliminate meaningless high frequency components in the calculation of cross-phase spectrum $R_{FG}(k_1, k_2)$ depending on the frequency spectrum of the given wood images~~-~~. When the 2D IFT is applied ~~on~~into the Equation (3.5), the POC function is generated as follows.

$$r_{fg}(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1,k_2} R_{FG}(k_1, k_2) \qquad (3.5)$$

Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)

Where $\sum k_1, k_2$ denotes$\sum_{n_1}^{M_1} = -M_1, \sum_{n_2}^{M_2} = -M_2$. When two images are similar, their POC function gives a distinct sharp peak. ~~When~~Similarly, when two images are not similar, the peak drops significantly. The height of the peak gives a good similarity measure for ~~images~~image matching, and the location of the peak ~~show~~shows the translation displacement between the images (Ito et al., 2008, Arnia & Pramita, 2011).

3.6 FEATURE ANALYSIS

Feature analysis is the final phase of the wood species recognition system. This phase involves the process of ~~analyzing~~analysing the characteristics of the resulting image after the feature extraction process. The aim is to find every kind of species tested with precise image using techniques that have been identified.

3.6.1 K~~–~~-Nearest ~~Neighbors~~Neighbours

In pattern recognition, the ~~k~~K-nearest neighbour algorithm (~~K-NN~~KNN) is generally used for classifying objects based on closest training examples in the feature space. K~~–~~-nearest ~~neighbors~~neighbours is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (Hassan et al., 2008).

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

~~K-NN~~KNN is a type of instance-based learning where the function is only approximated locally and all computation is deferred until classification. Training process for this algorithm only consists of storing feature vectors and labels of the training images. In the classification process, the unlabelled query point is simply assigned to the label of its k nearest ~~neighbors~~neighbours.

The idea in ~~k-Nearest Neighbor methods~~KNN is to identify k samples in the training set whose independent variables x_r are similar to x_s, and to use these k samples to classify this new sample into a class, v. If all we are prepared to assume is that f is a smooth function, a reasonable idea is to look for samples in our training data that are near it and then to compute v from the values of y for these samples. Before we get into the details of the ~~K-NN~~KNN algorithm we need to define the Euclidean distance between two input vectors x_r and x_s.

Let's consider the two input variable case since it is easy to represent in two-dimensional space. See the two vectors $x_r = (x_{r1}, x_{r2})$ and $x_s = (x_{s1}, x_{s2})$ represented in Figure 3.19. The distance between these two vectors is computed as the length of the difference vector, denoted by

$$d(x_r, x_s) = |x_{r1}, x_{s1}| = \sqrt{(x_{r1}, x_{s1})2 + (x_{r1}, x_{s1})2} \qquad (3.6)$$

The simplest case is $k = 1$ where the sample in the training set that is closest to u and set $v = y$ where y is the class of the nearest ~~neighboring~~neighbouring sample.

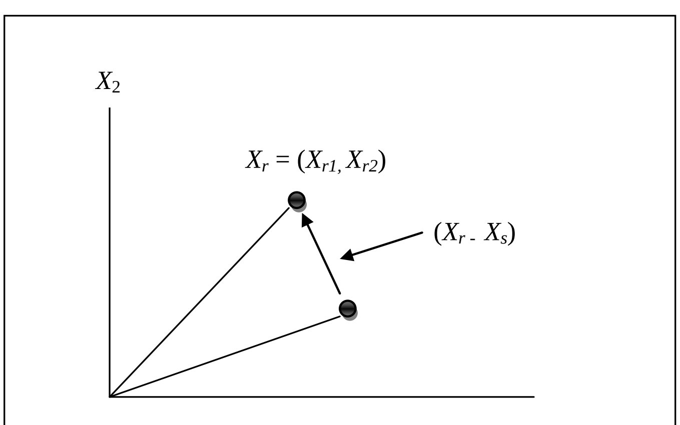
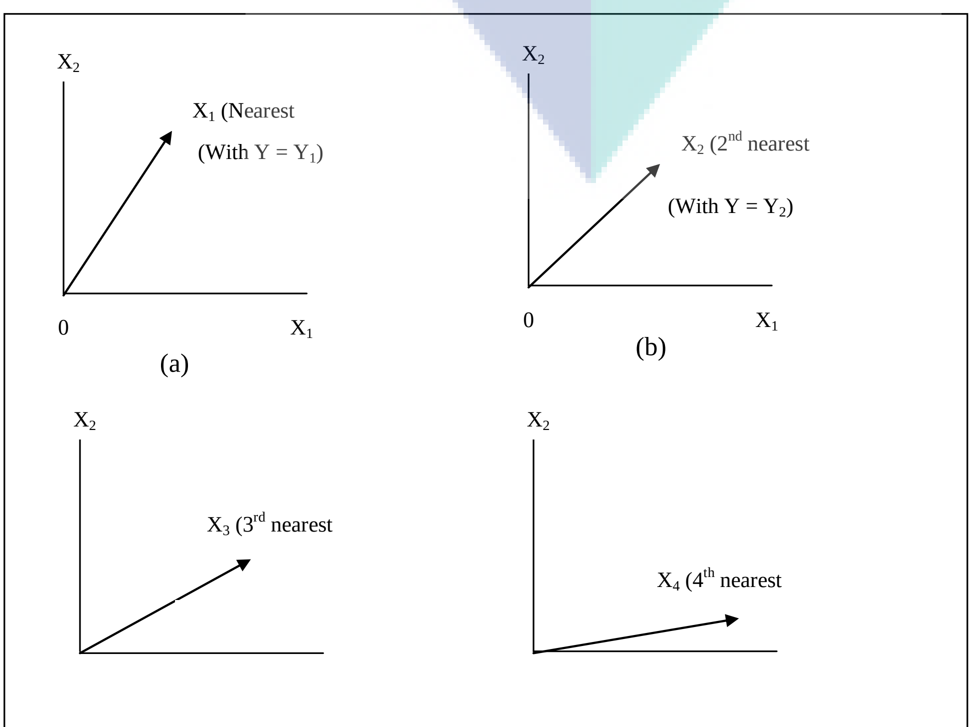


Figure 3.19 : The Euclidean Distance between two vectors x_r and x_s

To get started on how the ~~K-NN~~KNN method works, consider the representation in Figure 3.20. ~~On the~~The left-hand-side~~ are~~ represented four separate points in the training data set, namely,

$$X_1 = (x_{11}, x_{12}),\ X_2 = (x_{21}, x_{22}),\ X_3 = (x_{31}, x_{32}),\ X_4 = (x_{41}, x_{42}) \qquad (3.7)$$

~~with~~With their associated output values, respectively, y_1, y_2, y_3 and y_4. Figure 3.20 (a), (b), (c) and (d) ~~shown~~show the first input vector in the validation data set $x_0 = (x_{01}, x_{02})$. That is the wanted~~ to use a~~ method to predict the associated output value, y_0.



Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Bold, English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Italic, English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Italic, English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Italic, English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Italic, English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Italic, English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Italic, English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Italic, English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Italic, English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Italic, English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Italic, English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)

Figure 3.20: The feature vector x_0 using some of its ~~neighbors~~neighbours in the training data set

~~In~~ Figure 3.20 ~~show~~shows the positioned x_0 so that its closest ~~neighbor~~neighbour is x_1, that is distance $d_1 = d(x_0,x_1)$ from x_0. The ~~nest~~next closest ~~neighbor~~neighbour is x_2 that is distance $d_2 > d_1$ and ~~the~~ similarly for other points x_3 and x_4, in training data set such $d_4 > d_3 > d_2 > d_1$. The propose ~~algorithm~~algorithms used for classification are shown in ~~figure~~Figure 3.21 respectively.

Procedure: Classification

Input : Unknown wood species and the feature database

Output : The class of the wood species to which this unknown wood species is assigned

Begin Testing

1. Given unknown wood species.
2. Calculate KNN the distance of feature vector of the test image and feature vectors in training image.
2. Compare KNN value with test image and training image.

End Testing

Figure 3.21: Classification algorithm

Formatted: English (U.K.)

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

3.7 SUMMARY

~~As the~~In conclusion, ~~chapter~~Chapter 3 ~~is~~ explained ~~about~~ the methodology of the wood species recognition system. In this system, there are four (4) phases included. They are image acquisition phase, pre-processing phase, feature extraction phase and feature analysis phase.

Each phase has its own objectives and interests. The most important phase in order to ~~recognize~~recognise the wood species is feature extraction phase, where ~~this phase is implemented~~ the Phase-only Correlation (POC) technique is implemented during the process. During this phase, the wood is ~~recognized~~recognised and then it is classified according to the species. The classification process is done with the help of the K-Neighbour Network (KNN)~~–~~ technique where it helps to classify the wood images into its classes.

The POC involves three (3) steps during the feature extraction phase. The steps are the 2D Discrete Fourier Transform (2D DFT~~) step,~~), Cross Phase Spectrum ~~step~~ and 2D Inverse Fourier Transform (2D IFT~~) step.~~). Each step contains specific calculation and algorithm in order to get details on woods features and match it with the possible matching species. By applying all the three (3) steps, the testing image is found to be a match or mismatch with the training data. Then, KNN takes place ~~on~~in the process.

The results ~~on each phase are discussed on the next chapter. The chapter is discussed on the detail results~~ obtained from the experiments and algorithms used on each phase~~.~~ are discussed in the next chapter.

CHAPTER 4

RESULT AND DISCUSSION

4.1 INTRODUCTION

Once the prototype is developed, all eight (8) Malaysian wood species identified were tested using four (4) main phases for the wood recognition species system. Each phase in this system ~~–~~ produces its own output. This chapter will discuss in detail the result for each phase including the outputs and the analysis of the outputs.

4.2 RESULTS

For the result of this system, it has a main page ~~which~~that is the interface of the system. Actually, while using the MATLAB software, there is only one interface in the system and all the processes are done in the same interface but with different functionality. For this system, the recognition process is done step by step from image training till image testing to obtain the result. ~~Through this~~The system~~, the~~ is user ~~just has~~friendly, as users only have to click the buttons in the system. The recognition and classification of the wood species process started with the entering ~~the~~of training images and testing images. The testing image is the image that the user wants to test and it ~~displays~~is displayed in the box ~~that is~~ provided in the system.

4.2.1 Acquisition Phase

The row of data for the research is taken from the Malaysian Timber Industry Board (MTIB). The database consists of 1,280 (in one thousand two hundred and eighty rows) wood images with 160 images per wood species. Out of 1,280 images, 640 images were used for training ~~and~~ while the remaining 640 were used for testing purpose. In Figure 4.1 ~~sample shown,~~, it shows the sample of wood images in the database. The full sample as shown (see *Appendix A*).

Figure 4.1: Sample database

Formatted: English (U.K.)

4.2.2 Pre - processing Phase

During the pre-processing phase, the input image from the acquisition phase is taken. The input ~~color~~ colour images are then processed for ~~grayscale transformation processes. In~~ grey scale transformation ~~process,~~, where the colour image is converted into grey scale image. After the ~~grayscale~~grey scale transformation process, the images then ~~have undergone another process which is image enhancement process. For~~undergo the image enhancement ~~process, noise is~~ phase, which removed ~~from~~ the ~~image.~~ noise.

4.2.2.1 Result of ~~Gray Scale~~Greyscale Transformation

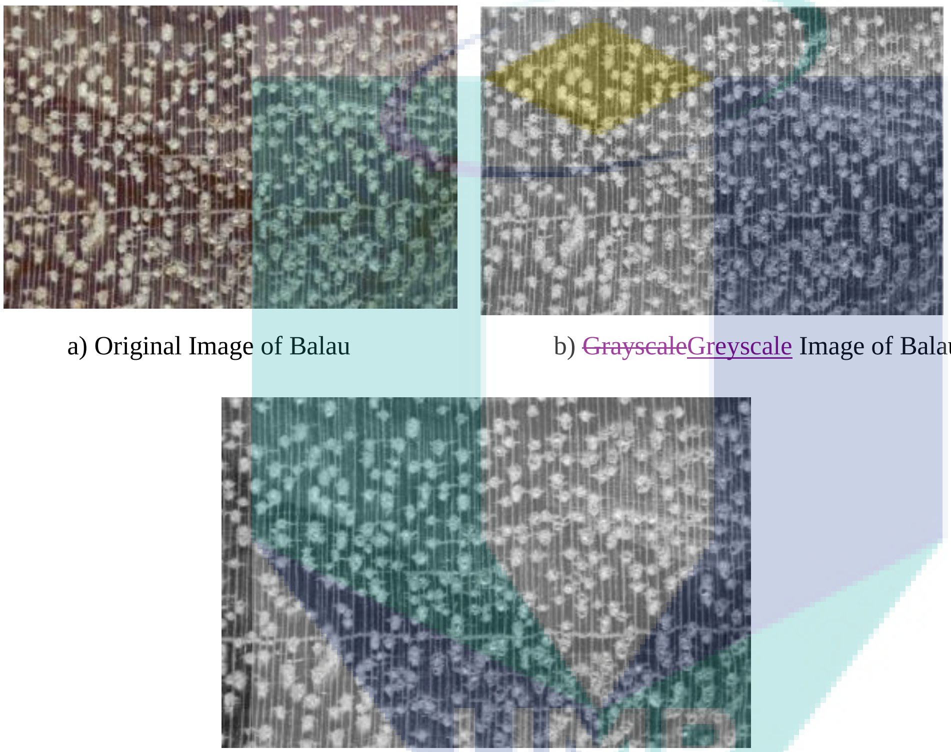
A ~~gray~~grey scale digital image is an image in which that the value of each pixel is ~~as~~a single sample, that is, it carries only intensity information. Images of this sort, also known as black-and-white, which are composed exclusively of shades of ~~gray, are varying~~grey, varied from black at the weakest intensity to white as the strongest intensity.

~~Gray~~Grey scale images contain the entire pixel values R, G, B ~~values~~ to be the same. The codes for the transformation are default code from the MATLAB software tools. The result conversion of some of the given wood images into ~~gray~~grey images is shown in Figure 4.2. The full sample shown (see *Appendix B*).

4.2.2.2 Result of Image Enhancement

After the grey scale transformation process, the wood species are then enhanced by ~~the~~ average filtering ~~for~~of the image enhancement process. The aim of this process is to remove the noise from the images. Figure 4.2 shows the results of the process~~.~~ ~~From the results, we know~~, which indicated that the average filtering technique ~~gives~~offers significant images with less noise found in the images. The full sample as shown (see *Appendix B*).

Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)



a) Original Image of Balau

b) ~~Grayscale~~Greyscale Image of Balau

c) Average Filtering Image of Balau

Figure 4.2: Result of image enhancement: (a) original image, (b) ~~grayscale~~grey scale image, (c) average filtering image

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

4.2.3 Result of Feature Extraction Phase

This is the third main phase ~~in~~ of this system. In this phase, the features of the woods are extracted. This is ~~an~~ important ~~phase since~~as its effective functioning improves ~~the~~ recognition rate and reduces ~~the~~ misclassification. This phase only involves ~~phase only~~ the correlation technique, which ~~include~~includes three steps ~~of~~namely the ~~technique. They are~~ 2D Discrete Fourier Transform (2D DFT), Cross-~~-~~phase spectrum and 2D Inverse Fourier transform (2D IFT).

During the 2D DFT step, the images are transformed into 256 bins where the frequency of each bin is calculated and then the mean frequency of each nine (9) consecutive bin is calculated. In this step, all the ~~images~~ training and testing images are involved.

Then, the cross phase spectrum took over the process. The cross phase spectrum transforms the data obtained from the previous step into a line graph. During this step, two (2) images from training images and testing images are compared. The aim of this step is to form a specific pattern of ~~the~~ image and compare it to another image. The data from 29 mean frequencies is converted into a pattern for each image.

Figure 4.3 ~~till~~to Figure 4.10 show the sample of ~~the~~ pattern for each wood species where it is defined from the data obtained from 2D DFT (Step 1). Figure 4.3 shows the sample pattern of the Balau species. This sample is taken from the training images. From Figure 4.3 ~~till figure~~to Figure 4.10, it is found that the pattern of each species and each image is genuinely different.

Figure 4.3: Pattern of the Ramin species training image

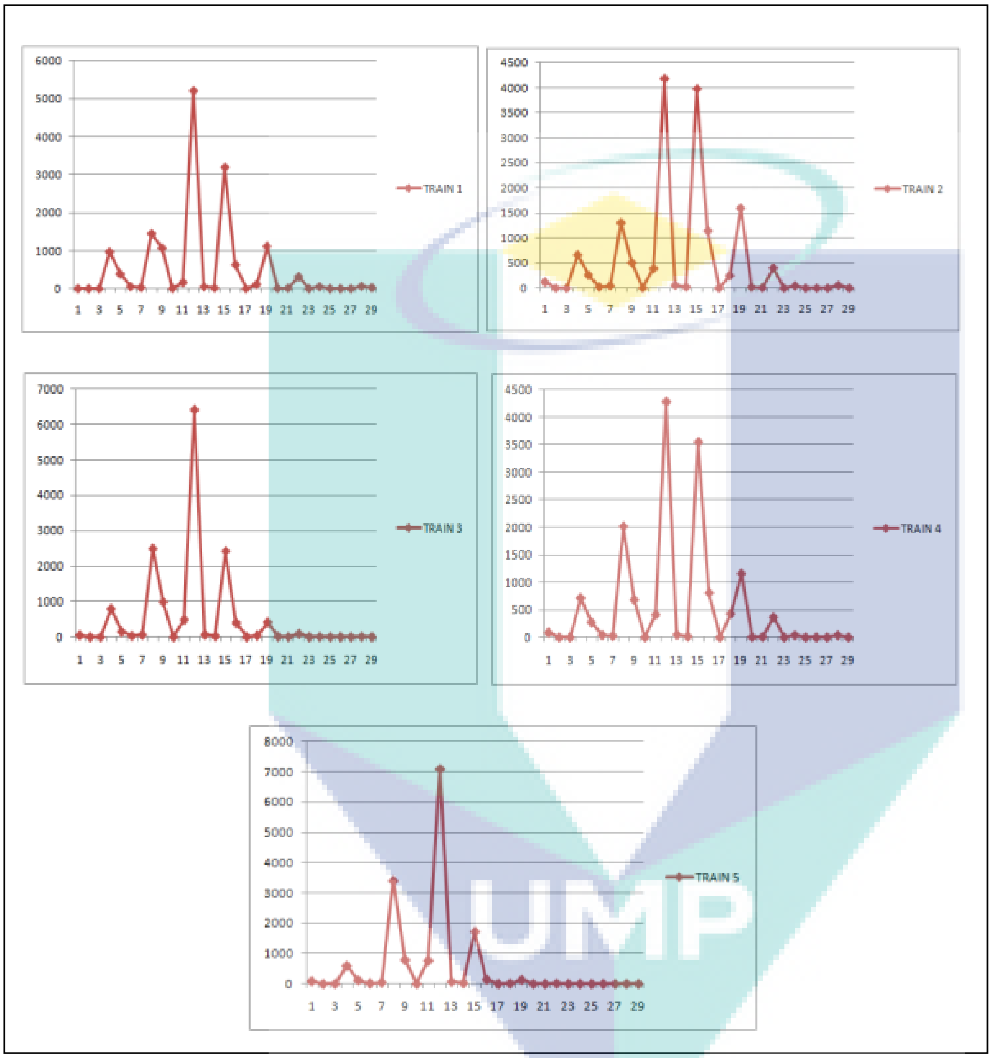


Figure 4.4: Pattern of the Melunak species training image

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Figure 4.5: Pattern of the Mersawa species training images

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Figure 4.6: Pattern of the Durian species training images

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

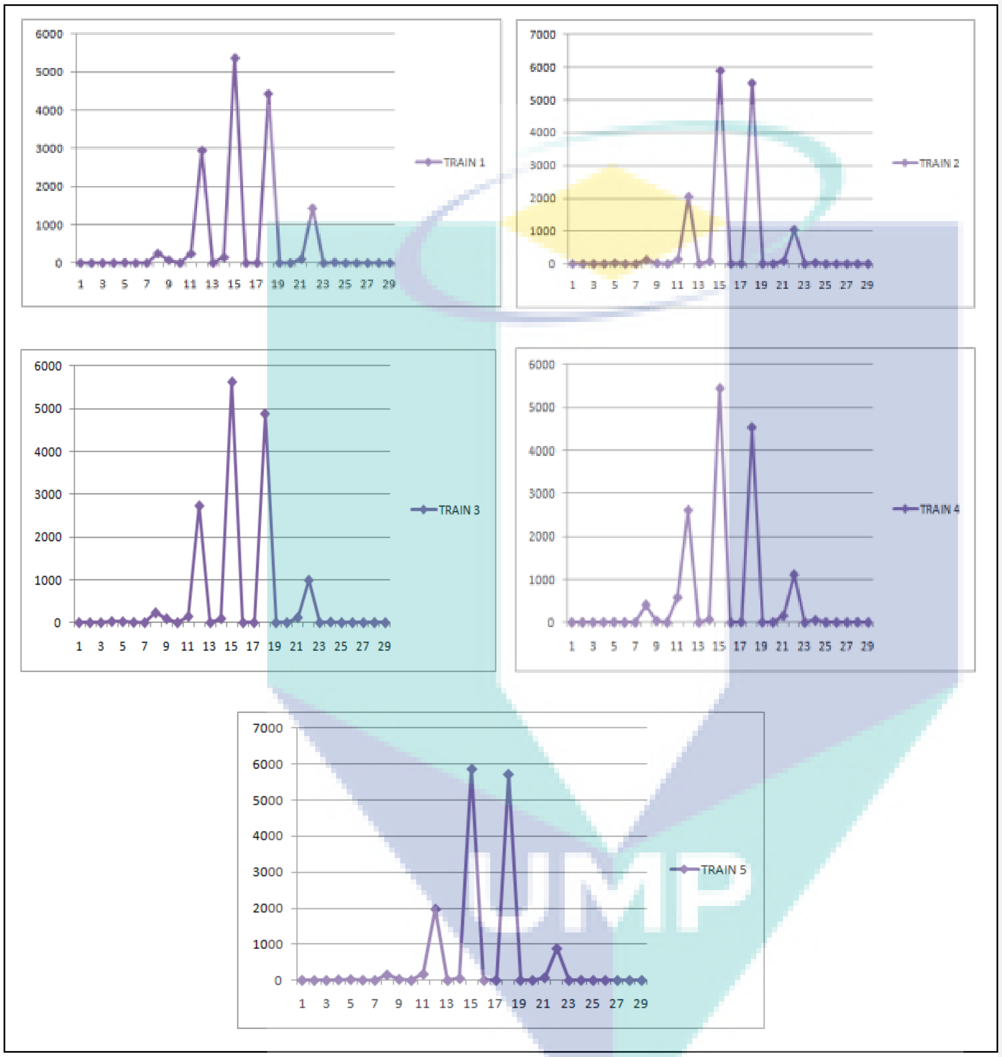


Figure 4.7: Pattern of the Mersawa Gajah species training images

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

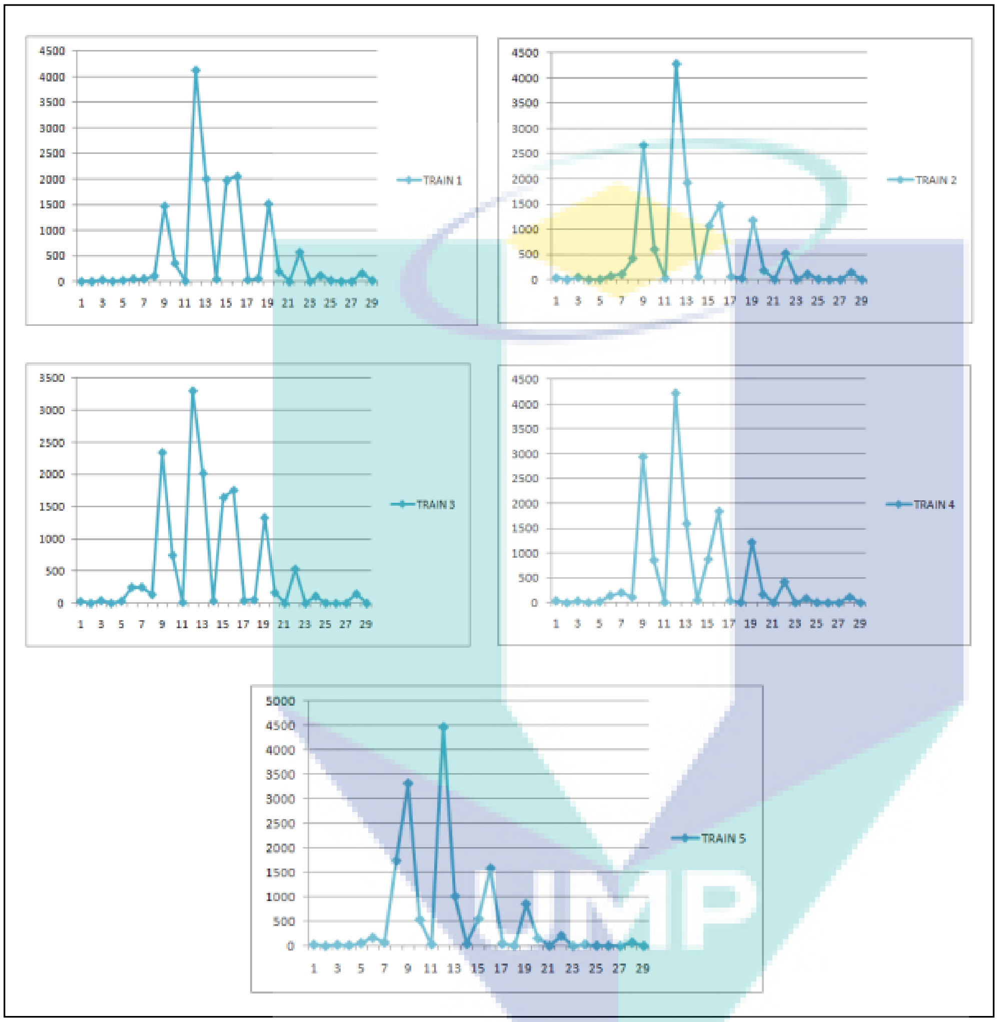


Figure 4.8: Pattern of the Rengas species training images

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

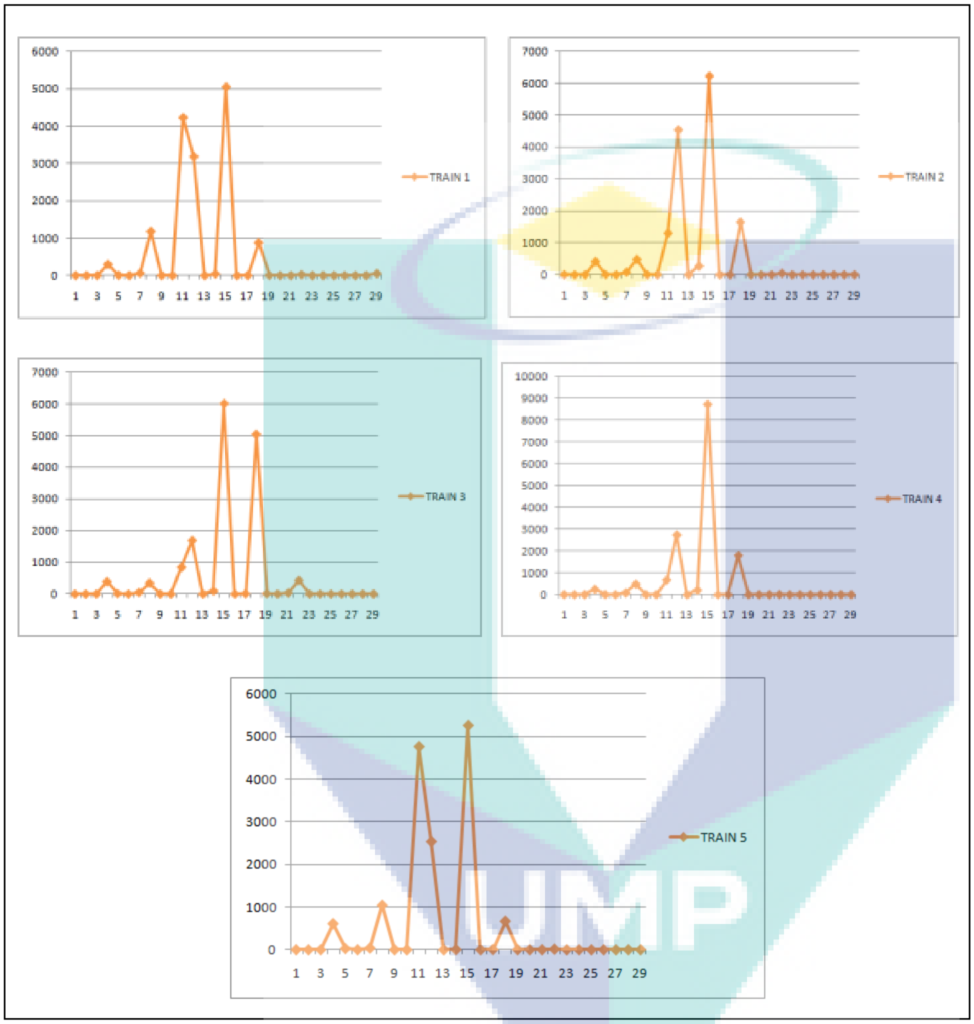


Figure 4.9: Pattern of the Sepetir species training images

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Figure 4.10: Pattern of the Balau species training images

Formatted: Font: Bold, English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)

When ~~pattern~~ the patterns of two (2) images are found to be matched, ~~the~~ a third step is involved. The 2D IFT then compared the two (2) patterns and find the closest match value of the two (2) images and performed the algorithm. Through the algorithm, the unrelated frequencies of the two (2) images are then eliminated and the distinct high peak frequency is then formed.

When two (2) images are compared and after the third step is performed, ~~during this step,~~ the 2D IFT eliminates all meaningless high mean frequencies obtained from the cross phase spectrum ~~step.~~ phase. It leads to the ~~existences~~ existence of a distinct sharp peak if the two (2) images (training image and testing image) are found to be the same. Meanwhile if they are dissimilar, the peak drops significantly.

Figure 4.11 shows two (2) images of the Durian species that are compared and ~~they are~~ found similar, so the POC function gives a distinct sharp peak. On the other hand, Figure 4.12 shows no distinct peak found when the Ramin species image (testing image) is compared to the Durian species image (training image). The height of the top peak is the good measurement to judge the similarity between two (2) images. Thus, the POC function exhibits much higher discrimination capability than ordinary correlation function.

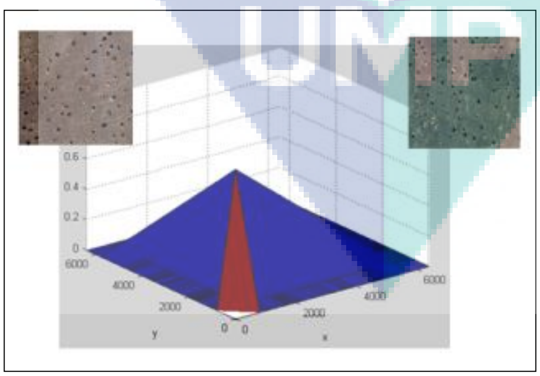


Figure 4.11: POC Function distinct sharp peak at centre for the same image types

Formatted: English (U.K.)

Formatted: Font: Bold, English (U.K.)

Formatted: English (U.K.)

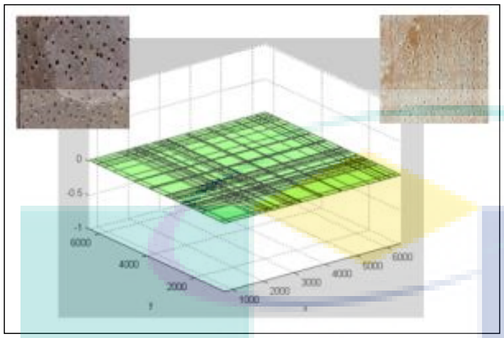


Figure 4.12: POC Function - no distinct peak for two ~~difference~~(2) different images

4.2.4 Feature Analysis Phase

Feature analysis is the last phase of the wood species recognition system where it prepares the output of the system. This phase used the K-~~nearest neighbour~~Nearest Neighbour (KNN) technique in order to classify the images according to their species. The analysis phase is the decision making part of a recognition system and it uses the data obtained from the features extraction in the previous phase.

Table 4.1 shows the classification of wood species using the K-Nearest Neighbour (KNN~~). This~~) technique, which is normally used for classifying objects. KNN is used after the POC process where the process involved the mean frequency of the vector value for each wood species. It computed the vector values of each wood species and found the nearest or minimum differences of the vector values between the data and the matched image.

KNN ~~is worked~~works when ~~the~~ two (2) images are matched after the process of POC during the previous phase. The KNN compares the vector value of the matched image with the vector value of all wood species. During the process, the minimum or nearest differences between them is found.

Formatted: Font: Bold, English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: Font: Bold, English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)
Formatted: English (U.K.)

Table 4.1: Wood species classification

Vector Value	Class	Wood Species
4190	C_1	Balau
8100	C_2	Ramin
5560	C_3	Durian
5440	C_4	Melunak
5420	C_5	Mersawa
5740	C_6	Mersawa Gajah
4080	C_7	Rengas
6220	C_8	Sepetir

From Table 4.1, it is found that the Balau's vector value is 4190~~. While~~ while for ~~rengas~~Rengas and Sepetir, the vector values are 4080 and 6220~~.~~ respectively. For Mersawa Gajah, Durian and Mersawa, the values are each 5740, 5560 and 5240.

The vector value for Melunak and Ramin, on the ~~vector values~~other hand, are 5440 and 8100. During the KNN process, the vector value of the matched image and all the vector values of all species are plotted on a graph. Then the differences between the value of all species and the matched image ~~is~~are calculated. The nearest or minimum differences between the values are considered as the same class.

For example, when two (2) images are considered as matched in the POC process, ~~and~~ its vector value is 6213. The KNN computes the ~~different~~difference of this image's vector value with all the species. After the calculation, it is found that the minimum ~~differences~~ difference is with the Sepetir wood species with ~~about~~ only seven (7) differences ~~rather than with~~compared to the Mersawa Gajah with almost 473 differences. Then the matched image is classified as the Sepetir wood species.

But when the vector value of the matched image is larger than 9100 vector value and smaller than 3080, the KNN cannot compute the minimum differences and the matched image is classified as ~~unrecognized. Besides that~~unrecognised. In addition, when the minimum differences of two (2) species are the same or two (2) wood species share the same minimum differences, the matched image is also classified as ~~un recognized~~unrecognised.

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

For example, the matched image's vector value is 5490. After the KNN computed the differences between all the species, it is found that Mersawa and Mersawa Gajah share the same differences ~~with~~at 250~~ differences. Then, the~~. The matched image is then considered as ~~unrecognized~~unrecognised.

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

Formatted: English (U.K.)

4.3 RESULT FOR IMAGE TESTING EXPERIMENT

From all the ~~result~~results obtained from each phase~~ included~~, there are certain elements valued in order to show the effectiveness of the system and whether the system is able to achieve the objectives of the research.

For each experiment, the research used the percentage of correct classification (PCC) to evaluate the classification accuracy, which is defined as:

$$PCC = \frac{number\ of\ testing\ images\ correctly\ classified}{total\ number\ of\ testing\ images\ classified} \times 100 \qquad (4.1)$$

The results of the experiments are ~~summarized~~summarised in ~~Tables~~Table 4.2 and Figure 4.11 with each wood species ~~use~~using the same 640 training data and are exactly evaluated on the same amount of testing data.

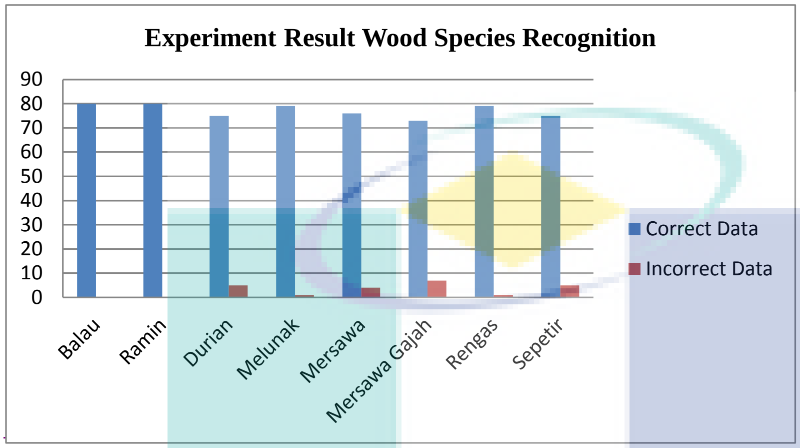
Table 4.2: Experiments result of wood species recognition

Formatted: English (U.K.)

Wood Species	Testing Data	Correct Result	Incorrect Result	Percentage (%)
Balau	80	80	0	100
Ramin	80	80	0	100
Durian	80	75	5	93.75
Melunak	80	79	1	98.5
Mersawa	80	76	4	95
Mersawa Gajah	80	73	7	91.25
Rengas	80	79	1	98.5
Sepetir	80	75	5	93.75

Based on Table 4.2, ~~there~~ about eighty (80) sample images for each species were tested. The experiment shows that all samples tested for the Balau and Ramin species are correctly identified. Meanwhile, for the Melunak and Rengas species, there is only one (1) incorrect image data ~~is~~ identified.

Furthermore, the Mersawa species obtained four (4) identified incorrect image data ~~identified. Otherwise~~. Likewise, both the Durian and Sepetir species are incorrectly identified for five (5) image data. Lastly, the highest incorrect image data identified is the Sepetir species with seven (7~~ image data~~) unsuccessful ~~identified.~~image data. The percentage for the results of the wood sample test shows that the recognition accuracy obtained is 96.4%.



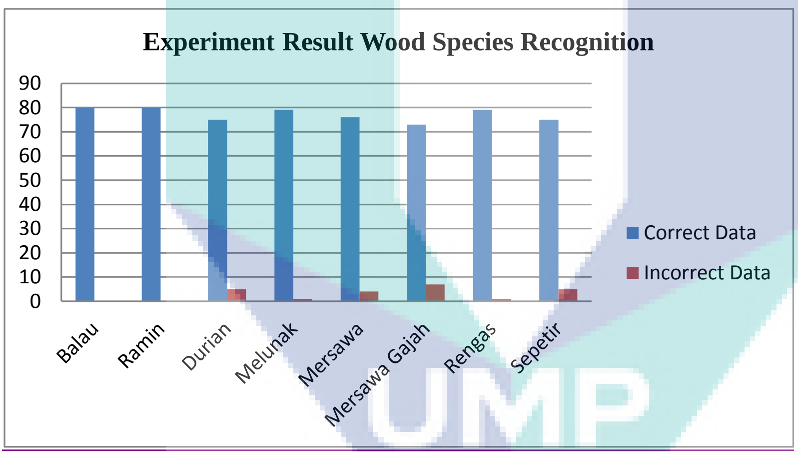


Figure 4.13: Histogram of the Wood Species Recognition Result

Formatted: English (U.K.)

Formatted: English (U.K.)

~~Therefore, based~~Based on Table 4.2, the histogram of the overall result wood species image data, which has been showed in Figure 4.11, has been plotted to ~~summarize~~summarise the table. From the histogram, the result illustrates that the Balau and Ramin species have ~~a~~the same result with all eighty (80) correct image data~~ is correct~~.

Meanwhile, for the Melunak and Rengas species, there are 79 image data ~~are~~ correctly identified. ~~Furthermore, the result~~As for the Mersawa species, it obtained 76 image data identified correctly~~. Otherwise,~~ while both the Durian and Sepetir species are correctly identified for 75 image data. Lastly, the lowest correct image data identified is the Sepetir species with only 73 image data ~~are~~ identified correctly.

4.4 COMPARISON AMONG RESEARCHES

Table 4.3: Comparison among researches in wood species recognition system

Wood Species Recognition System By (Bremananth et al. 2009)	Design of an Intelligent Wood Species Recognition System By (Khalid et al. 2008)	Wood Species Recognition System Based on Phase Only-Correlation (POC)
Using GLCM approach	Using GLCM and Multilayer Perceptron Artification Neural Network	Using Bin Based Histogram and POC
Involves 2 steps: 1. Pre-processing 2. Feature Extraction	Involves 2 steps: 1. Pre-processing 2. Feature Extraction	Involves 3 steps: 1. Pre-processing 2. Feature Extraction 3. Feature Analysis
200 train images 200 test images	1753 train images 196 test images	640 train images 640 test images
Accuracy percentage: Not stated	Accuracy percentage: 95%	Accuracy percentage: 96.4%

According to ~~the~~ Table 4.3 above, this research is compared to two (2 ~~researches~~) separate studies in wood species recognition field. They are Wood Species Recognition System by Bremananth R et al. and Design of an Intelligent Wood Species Recognition System by Khalid Marzuki et al. ~~Those both systems~~Both the researches used GLCM as the ~~basic~~basis of their system.

For Wood Species Recognition System (by Bremananth et al. 2009)~~ system,~~, he implemented his research on ten (10) Indian woods species ~~which are~~including Teak, Ebony, Jack, Oak, Padauk, Sal, Satin, Teak, White Oak and Zebra. He used only the GLCM approach and ~~involves~~involved two (2) steps during the process. The steps ~~are~~were pre-processing and feature extraction. During the pre-processing step, the images obtained ~~are~~were resized to 256 x 256 and the images ~~are~~were converted into ~~the gray~~grey images. For the feature extraction, after considering ~~the~~ several textural analysis methods, the ~~Gray~~Grey Level Co-occurrence Matrix (GLCM) ~~seems~~was deemed as appropriate.

In ~~other ways~~contrast, the second research, (Khalid et al. 2008) implemented their research on twenty (20) Malaysian woods species, which are Bintangor, Nyatoh, Sesendok, Ramin, Mersawa, Jelutong and others. They used the GLCM and Multilayer Perceptron (MLP) Artificial Neural Network (ANN). It ~~involves~~ involved two (2) steps during the process, which ~~are~~were pre-processing and feature extraction.

In the pre-processing step, the image enhancement used two procedures of the Visual System Development Platform (VSDP) Image Library. They ~~are~~were image sharpening using high-pass filter~~ and~~, as well as contrast enhancement using histogram equalization. ~~While for~~For the feature extraction step, he used GLCM as the feature extractor and Multilayer Perceptron (MLP) Artificial Neural Network (ANN) as the wood species classification. From the 20 Malaysian wood species, he used 1753 ~~train~~training images and 196 test images. As ~~the~~a result, the accuracy percentage obtained ~~is~~was 95%.

~~Different with~~ Compared to this research, the ~~research~~researcher used eight (8) Malaysian woods species to implement the POC method. This research involved three (3) steps during the process, which ~~are~~were pre-~~-~~processing, feature extraction, and feature analysis. This research also enhanced the POC technique by adding the KNN ~~technique~~method in order to attain better results compared to other ~~research.~~researches. As

Formatted: English (U.K.), Kern at 12 pt

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

~~the~~a result, the accuracy percentage obtained ~~is~~amounted to 96.4% and the researcher believed that this advanced POC technique attained better results ~~from~~than other technique.

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

Formatted: Font: English (U.K.)

4.5 SUMMARY

This research ~~contains~~comprised of three (3) phases ~~which are~~, including pre-processing, feature extraction and feature analysis. In the pre-~~-~~processing phase, there are two (2) processes involved ~~which are~~, namely the grey scale transformation process and image enhancement process. In the grey scale transformation process, the colour image is converted into grey scale image. While for the image enhancement process, image noise is removed ~~from the image~~ by using the average filtering technique.

The next phase is feature extraction where the enhanced ~~grayscale~~grey scale image is computed by the phase only correlation (POC) technique. During this phase, the frequency of each bin is calculated and the sharp peak formed meant that the test image is matched with the train image.

In order to achieve the objectives of this research, the last phase is performed. In this last phase, the K ~~nearest neighbor~~ Nearest Neighbour (KNN) technique is implemented. The objective of this phase is to identify the name of the matched test image. When the POC found that the test image is matched with the train image, the KNN helped to identify the wood species name ~~using~~by calculating the feature vector values ~~calculated~~.

~~As the~~In conclusion, the incorrect results happened because of some factors and the first factor is the lighting problem. During the photography session, the light factor needs to be considered ~~so~~to ensure that the image ~~that has been~~ taken is not too dark or ~~not~~ too light. If this happened, then it causes difficulty in deciding the wood species. Next~~,~~ is the way ~~of the~~ images are taken. ~~Actually, there~~There are ~~a few~~several steps that must be followed during the photography session.

The camera position must not be too ~~far~~ near or too ~~close to~~far from the image. ~~Besides~~In addition, the position should always be maintained and static for the next sample. ~~In addition, the~~ The camera must also ~~must~~ be placed ~~face~~facing towards the object ~~on~~of the image and not ~~slanting~~be slanted away from the image.

The next factor is ~~a~~the technique of cross section cutting. The rough or uneven cross section cutting might caused problems while testing the image. In fact, this condition ~~may caused the~~might result in error or incorrect ~~of the~~ data identification.

CHAPTER 5

CONCLUSION

5.1 INTRODUCTION

The aim of this study is to determine the most suitable methodology and techniques for ~~the~~ wood species recognition. In this research there are four (4) main phases used ~~which are the~~ image acquisition~~ phase, the~~, pre~~-~~-processing~~ phase~~, feature extraction ~~phase~~ and feature analysis~~ phase~~. Based on the research questions and results discussed in Chapter 4, ~~few~~several conclusions have been made and they are as ~~follows~~follow:

5.2 OBJECTIVES ACHIEVEMENT

In order to achieve the objectives of the research, there are certain phases that should take ~~places.~~place. The first ~~phase~~ is image acquisition~~. The image is~~, where images are collected from ~~the~~ data ~~of~~ provided by the Malaysian Timber Industry Board (MTIB). There are 1280 images consist of eight (8) wood species collected, with 160 images ~~collected~~ for ~~a~~each wood species. During~~ the~~ image capturing, the researcher used an ordinary digital camera. When the image is captured, the image is converted from an ~~analog image~~analogue to a digital image.

The second phase is pre-processing~~. The aim of this phase~~, which is aimed to enhance the image. During this step, the colour image is converted into ~~the grayscale image. Besides that~~grey scale. In addition, the noise is removed by using the average filtering technique.

~~After that,~~ Next is the third phase ~~is took place. The step~~, which is feature extraction. ~~During the feature extraction~~ In this phase, histogram and the Phase Only Correlation (POC) method is used to determine the wood species pattern. Histogram is used for processing to define the frequency of each ~~gray~~grey level present in the image while the POC ~~will~~would value between the test image and the image of predicted class. At this phase, the image is ~~recognized~~recognised whether it is a match or mismatch.

The ~~next~~subsequent phase is feature analysis~~. During this phase,~~ , where the K-Nearest Neighbour technique is implemented. The aim of this technique is to ~~recognize~~recognise the wood species and classify it by their species. The KNN ~~is functioning~~functions by calculating the feature vector of the testing image and compare it with the training image stored in the database. The minimum difference of the feature vector's value between the testing data and training data ~~is~~is considered as matched.

After completing all the ~~phase~~phases, the result of each phase is collected and presented. It is found that the percentage for the accuracy of the wood species recognition system based on the POC obtained is 96.4%. From ~~the~~ this percentage, it is shown that the system ~~is~~has fulfilled the objectives of this research.

5.3 CONTRIBUTION

The main ~~contributions~~contribution of this research ~~are the research is proposed in order~~ is to produce a new technique in enhancing the wood species recognition system based on the Phase Only–Correlation (POC) technique in contrast ~~of the~~ with previous researches ~~which~~that used the Grey Level Co-occurance Matrices (GLCM) technique.

~~Since~~This research is motivated by the unsuccessful results obtained from the previous ~~research found unsuccessful, it is motivated to propose this research.~~study. This new technique obtained good results with ~~high~~higher percentage of identification. It also achieves the most significant features in identifying ~~the~~ wood species.

In addition, in order to achieve a better result, this research enhances the POC process by applying the K-Nearest Neighbour technique. Based on the process of POC in the previous study on other field, the KNN technique is ~~decided~~selected to be applied into the POC process ~~of POC~~. By applying this technique, the wood species ~~can~~could be ~~recognized~~recognised according to their species.

~~Besides that~~Furthermore, this system ~~will~~would be able to replace ~~the~~ wood inspectors who usually take more time in ~~recognizing the~~recognising wood species by using conventional techniques. ~~By using~~The technique recommended by this~~, it~~ research takes only a short time to ~~recognize~~recognise the wood species.

5.4 FURTHER RESEARCH

~~From the research done, it shows that~~Nevertheless, this research has several weaknesses~~. Thus, there are~~, which require some improvement ~~is suggested to be done~~ to improvise the system so that the ~~result~~results obtained ~~will~~would be much better.

i. The technique used to ~~recognize~~recognise the Mersawa and Mersawa Gajah species, which ~~the~~ characteristics ~~are look like the same~~looked similar, should be thoroughly investigated. It is suggested ~~that the~~to have continuous research ~~is continuously done in order~~ to attain a better methodology and techniques.

ii. The testing is only tested on eight (8) wood species. It is ~~hoped that it will be done~~proposed to conduct testing on more wood species so that the efficiency of the wood species recognition system ~~can~~could be improved and it ~~may~~would reduce errors while applying the wood species recognition system.

iii. The process of capturing the wood species images ~~becomes so hard~~is a challenge because of the lack of a systematic system in capturing the wood images in Malaysia. It is suggested that ~~other researcher should~~in another research to develop a system to make this process easier. Besides that, various images in ~~difference~~different forms could be used to identify more ~~situation of the~~situations, so that recognition of wood species ~~to get~~can secure more impressive ~~result~~results.

iv. The effort in improvising the methodology and techniques in wood species recognition system should ~~be continued~~continue to obtain a better result in the future. There are a lot of techniques that had been applied nowadays and in the future, a new system ~~can~~could be developed by applying other techniques and algorithm to produce ~~the~~ a better result and to apply it with ~~a lot of~~more images ~~based on~~ for wood recognition ~~and wood~~ species.

This project can also be applied in the wood industry because it can help to ~~recognize~~recognise and classify wood species in a short time .The new developer can develop it using other technique and apply it to the wood industry.

Formatted: Indent: First line: 0 cm

5.5 SUMMARY

~~Thus~~This research has developed a visual inspection system for the recognition of wood ~~is developed.~~species. The system was objectively designed to be cost-effective and as a means to replace wood inspectors due to the difficulty in ~~recruiting them as workers which is rather laborious.~~recruitment. The system shows an ~~effectual visualization~~effective visualisation of the application. The image processing techniques are applied to improve the image excellence. In this design we have applied the POC approach to extract the features from the digital wood images.

As a result, the recognition process of this system ~~become smoothly rehearses because of the steps that used in this system.~~becomes smoother. Even though this system ~~could give~~has several advantages to the user, ~~but this system was~~it will still take time in the recognition process.

REFERENCES

Abdelrahman, A. and Hamid, O. 2012. Breast Ultrasound Images Enhancement Using Gray Level Co-Occurrence Matrices Quantizing Technique. *International Journal of Information Science*, 2(5): 60–64.

Abdi, H., and Williams, L. J. 2010. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4): 433-459.

Alginahi, Y. 2010. Preprocessing Techniques in Character Recognition, Character Recognition, Minoru Mori (Ed.), ISBN: 978-953-307-105-3, InTech.

Ariff, M. 2005. Wood-based industry deserves more attention. *Malaysian Institute of Economic Research (MIERScan) Report.*

Arnia ,F. and Pramita, N. 2011. Enhancement of Iris Recognition System Based on Phase Only Correlation, Telkomnika, 9(2): 387-394, ISSN: 1693-16930.

Arntzen, Charles J. 1994, Wood Propertis. *ed. Encyclopedia of Agricultural Science*. Orlando, FL: Academic Press. 4(3) : 549-561.

Ashoka, H. N., Manjaiah, D. H. and Bera, R. 2012. Feature Extraction Technique for Neural Network Based Pattern Recognition. *International Journal on Computer Science and Engineering*, *4*(3).

Baginski, J., Babuska, L. and Klimczewski, M., 2007. Development of Quality Management in Polish Wood Industry. *Cost E53 Conference - Quality Control for Wood and Wood Products*.Warsaw, Poland.

Barbu, T. 2006. An automatic unsupervised pattern recognition approach. *Proceedings of the Romanian Academy*, *7*(1): 73-78.

Bebis, G., Egbert, D. and Shah, M. 2003. Review of computer vision education. *Education, IEEE Transactions on*, *46*(1): 2-21.

Bezdek, J. C., and Pal, S. K. 1992. *Fuzzy models for pattern recognition* (Vol. 267). New York: IEEE press.

Bingham, E., and Hyvarinen, A. 2000. A fast fixed-point algorithm for independent component analysis of complex valued signals. *International journal of neural systems*, 10(01): 1-8.

Borg, I., and Groenen, P. J. 2005. Modern multidimensional scaling: Theory and applications. Springer Verlag.

Bouckaert, R. R. 2004. Naive Bayes classifiers that perform well with continuous variables. In *AI 2004: Advances in Artificial Intelligence*, pp. 1089-1094. Springer Berlin Heidelberg.

Bond, B., Assistant, P. & Hamner, P., 2011. *Wood Identification for Hardwood and Softwood Species Native to Tennessee*, The University of Tennessee Institute of Agriculture, U.S. Department of Agriculture.

Bremananth, R., Nithya, B., and Saipriya, R. 2009. Wood species recognition system. *World Academy of Science Engineering and Technology*, *52*(128): 873-879.

Buehlmann, U., Lihra, T., Rancourt, V., and Ait-Kadi, D. 2007. Detection capabilities of automated hardwood lumber defect-detection systems. *Forest products journal*, *57*(10): 51-57.

Bullinaria, J.A., 2004. Self Organizing Maps : Fundamentals. Introduction to Neural Networks : Lecture 16. University of Birmingham.

Cavalin, P., Oliveira, L. S., Koerich, A. L., and Britto, A. S. 2006, November. Wood defect detection using grayscale images and an optimized feature set. In *32nd Annual Conf. of the IEEE Industrial Electronics Society*, pp.3408–3412.

Comon, P. 1994. "Independent Component Analysis: a new concept?", *Signal Processing*, 36(3): 287–314.

Dash, M., and Liu, H. (1997). Feature selection for classification. *Intelligent data analysis*, 1(1-4): 131-156.

De Vos, M., De Lathauwer, L., Vergult, A., De Clercq, W., Van Paesschen, W., & Van Huffel, S. (2006). Jacobi iterations for spatially constrained independent component analysis. In *Proc. of the 7th international conference on methematics in signal processing (IMA 2007), Cirencester, England*, (pp. 2-5).

Fisher, R.A. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* , 7(2): 179-188.

Forestry Department, 2010. State of the World's Forests, FAO, report.

Gadkari, D. 2004. *Image Quality Analysis Using GLCM*. Master Thesis of Science in Modeling and Simulation, University of Central Florida Orlando, Florida.

Girolami, M., and He, C. 2003. Probability density estimation from optimally condensed data samples. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *25*(10): 1253-1264.

Gnanasivam, P., and Muttan, S. (2012). Estimation of Age Through Fingerprints Using Wavelet Transform and Singular Value Decomposition. *International Journal of Biometrics and Bioinformatics (IJBB)*, 6(2): 58-67.

Gonzalez, R. C., Woods, R. E., and Eddins, S. L. 2004. Digital Image Processing Using MATLAB, Prentice Hall, Upper Saddle River, NJ.

Guru, D S., Sharath, Y. H. and Manjunath S. 2010. Texture Features and KNN in Classification of Flower Images. *IJCA, Special Issue on "Recent Trends in Image Processing and Pattern Recognition" RTIPPR* (1):21-29.

Hassan, M. R., Hossain, M. M., Bailey, J., and Ramamohanarao, K. 2008. Improving k-nearest neighbour classification with distance functions based on receiver operating characteristics. *In Machine Learning and Knowledge Discovery in Databases*, pp. 489-504. Springer Berlin Heidelberg.

Hawalick, R. M. 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5): 786-804.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 20(8): 832-844.

Hodge, V. J., and Austin, J. 2005. A binary neural k-nearest neighbour technique. *Knowledge and Information Systems*, *8*(3): 276-291.

Honkela, T. 1997. Self-organizing maps in natural language processing PhD Thesis. Helsinki University of Technology.

Hush, D. R., and Horne, B. G. 1993. Progress in supervised neural networks. *Signal Processing Magazine, IEEE, 10*(1): 8-39.

Ito, K., Aoki, T., Kosuge, E., Kawamata, R., and Kashima, I. 2008, December. Medical image registration using phase-only correlation for distorted dental radiographs. *In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* (pp. 1-4). IEEE.

Jain, A. K., Duin, R. P. W., and Mao, J. 2000. Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1): 4-37.

Khalid, M., Lee, E. Yusof, R. and Nadaraj, M. 2008. Design of an intelligent wood species recognition system. *International Journal of Simulation System, Science and Technology*, 9(3): 9-19.

Klock, H., and Buhmann, J. M. 1998. Data visualization by multidimensional scaling: a deterministic annealing approach. *Pattern Recognition*, 33(4): 651-669.

Kotsiantis, S. B., Zaharakis, I. D., and Pintelas, P. E. 2007. Supervised machine learning: A review of classification techniques. *Frontiers in Artificial Intelligence and Applications*, 160, (3) : 249-268.

Lampinen, J., Laaksonen, J., and Oja, E. 1997. Neural network systems, techniques and applications in pattern recognition. Helsinki University of Technology.

Lew.Y. L. 2005, Design of an intelligent wood recognition system for the classification of tropical wood species, M. E. thesis, Universiti Teknologi Malaysia, .

Looney, C.G. 2003. Pattern recognition. CRC Press.

Malaysian Timber Industry Board, 2009. Maskayu, report.

Medlock, B., & Briscoe, T. (2007, June). Weakly supervised learning for hedge classification in scientific literature. In *Annual Meeting-Association For Computational Linguistics* ,45(1): 992-999.

Meijer, B. R. (1992). Rules and algorithms for the design of templates for template matching. In *Pattern Recognition, 1992. Vol. I. Conference A: Computer Vision and Applications, Proceedings., 11th IAPR International Conference on* (pp. 760-763). IEEE.

Menon, P.K.B., 2004. Structure and Identification of Malayan Woods. Revised by Ani Sulaiman & Lim Seng Choon, *Malayan Forest Records No. 25,* Forest Research Institute Malaysia

Michie, E.D., Spiegelhalter, D.J. and Taylor, C.C. 1994. Machine Learning, Neural and Statistical Classification, *Series in Artificial Intelligence*, *New York: Ellis Horwood*.

Musa, Z., Abdul Kadir, T. A., and Abu Bakar, R. 2007, September. Textile Web Defect Inspection by Feature Analysis Method. In *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on* (pp. 376-376). IEEE.

Nakajima, H., Kobayashi, K., and Higuchi, T. 2008. A palmprint recognition algorithm using phase-only correlation. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 91(4): 1023-1030.

Neumann, J., Schnörr, C., and Steidl, G. 2004. SVM-based feature selection by direct objective minimisation. In *Pattern Recognition, Proc. of 26th DAGM Symposium* (pp. 212-219). Springer Berlin Heidelberg.

Nilsson, N.J. (1965). Learning machines. New York: McGraw-Hill.

Nomura, S., Yamanaka, K., Shiose, T., Kawakami, H., and Katai, O. 2009. Morphological preprocessing method to thresholding degraded word images. Pattern Recognition Letters, 30(8): 729-744.

Pan, G. 2005. A Tutorial of Wavelet for Pattern Recognition. Taiwan University, Taipei, Taiwan, ROC.

Pascual, A.F., 2010. Advanced methods for dimensionality reduction and clustering: Laplacian Eigenmaps and Spectral Clustering. *Master Thesis*. University Autonoma de Madrid.

Pradeep, J., Srinivasan, E. and Himavathi, S., 2010. Diagonal Feature Extraction Based Handwritten Character System Using Neural Network. *International Journal of Computer Applications*, 8(9): 17-22.

Rao, K. R., and Hwang, J. J. 2010. *Fast Fourier transform: Algorithms and applications*. Springer.

Shinde, S. and Deshmukh, V. 2011. Implementation of Pattern Recognition Techniques and Overview of Its Application In Various Areas of Artificial Intelligence. *International Journal of Advances in Engineering and Technology, IJAET*, 1(4): 127–137.

Silva, T.C., Cupertino, T.H. and Zhao, L., 2011. High Level Classification for Pattern Recognition. *Graphics, Patterns and Images SIBGRAPI*, pp.344 – 351.

Takita, K., Sasaki, Y., Higuchi, T., and Kobayashi, K. (2003). High-accuracy subpixel image registration based on phase-only correlation. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 86(8): 1925-1934.

Tohka, J., 2011. SGN-2506: Introduction to Pattern Recognition.Institute of Signal Processing, Tampere University of Technology.

Tou, J. and Gonzalez, R.C. 1974, .Pattern Recognition Principles. Addison- Wesley Publishing Company .

Tou, J. Y., Tay, Y. H., and Lau, P. Y. 2008. One-dimensional grey-level co-occurrence matrices for texture classification. In *Information Technology, ITSim 2008. International Symposium* on (Vol. 3, pp. 1-6) IEEE.

Tou, J. Y., Lau, P. Y., and Tay, Y. H. 2007. Computer vision-based wood recognition system. In *Proceedings of International Workshop on Advanced Image Technology IWAIT*,Bangkok, pp.197-202.

Umbaugh, S. E. 1998. Computer Vision and Image Processing: A Practical Approach Using CVIPtools. Prentice-Hall International,Inc.

Verma, R. and Raja, J., 2005. Surface Defect Analysis Using Pattern Recognition Techniques. *American Society for Precision Engineering, University of North Carolina at Charlotte*. Metrology/Surface and Subsurface/Poster/Abstract/2005 Annual, pp.1–4.

Webb, A. R. 1995. Multidimensional scaling by iterative majorization using radial basis functions. *Pattern Recognition*, *28*(5): 753-759.

Watanabe, S. 1985, Pattern Recognition: Human and Mechanical. New York: Wiley.

Yang, Y., and Webb, G. I. 2003. On why discretization works for naive-bayes classifiers. *In AI 2003: Advances in Artificial Intelligence* (pp. 440-452). Springer Berlin Heidelberg.

Zhang, J., Ou, Z., and Wei, H. 2006. Fingerprint matching using phase-only correlation and fourier-mellin transforms. In *Intelligent Systems Design and Applications, 2006. ISDA'06. Sixth International Conference on* (Vol. 2, pp. 379-383). IEEE.

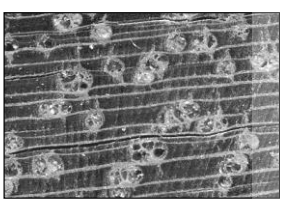
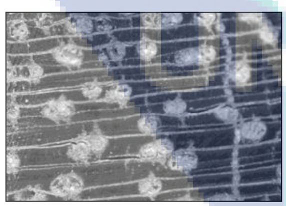
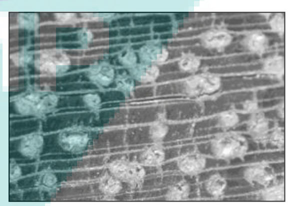
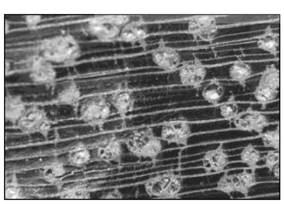
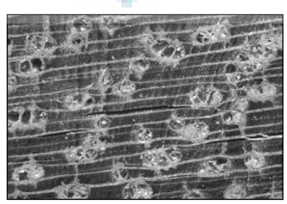
Zheng, L. & He, X., 2005. Classification Techniques in Pattern Recognition. In *WSCG'2005*. Plzen,Czech Republic: Science Press.

APPENDIX A

SAMPLE OF WOOD SPECIES IMAGE

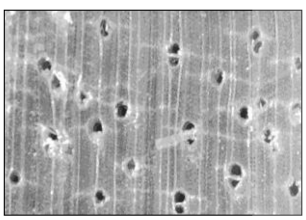
APPENDIX A1

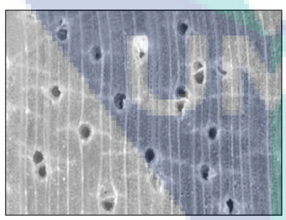
UMP DATABASE: 5 IMAGES OF BALAU WOOD SPECIES

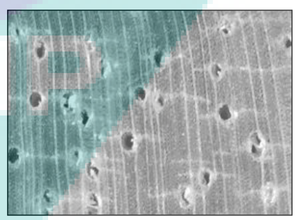






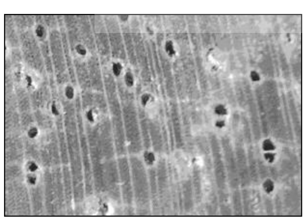
APPENDIX A2

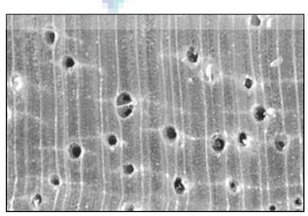
UMP DATABASE: 5 IMAGES OF RAMIN WOOD SPECIES





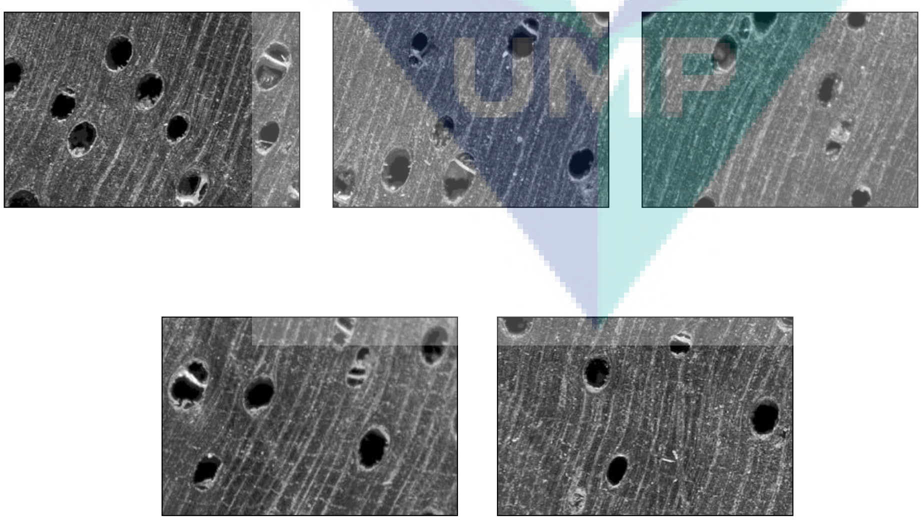






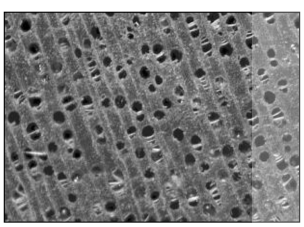
APPENDIX A3

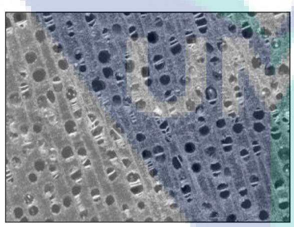
UMP DATABASE: 5 IMAGES OF DURIAN WOOD SPECIES

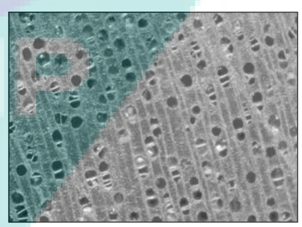


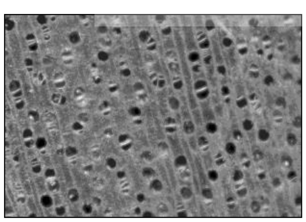
APPENDIX A4

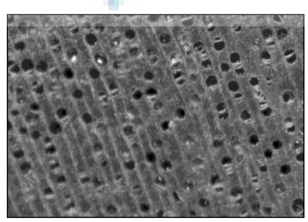
UMP DATABASE: 5 IMAGES OF MELUNAK WOOD SPECIES





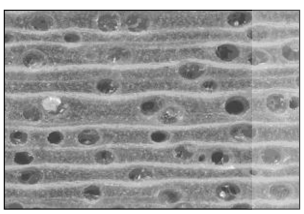
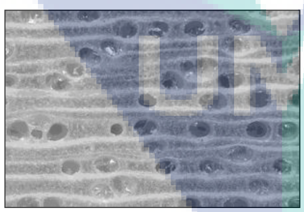
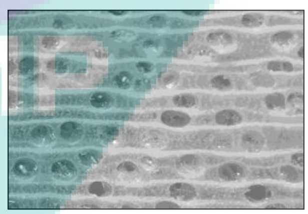
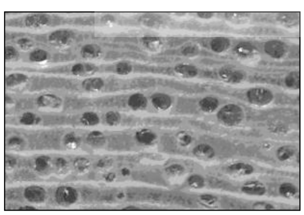
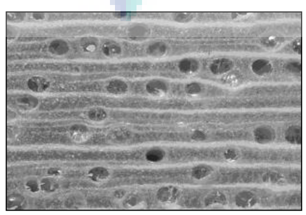






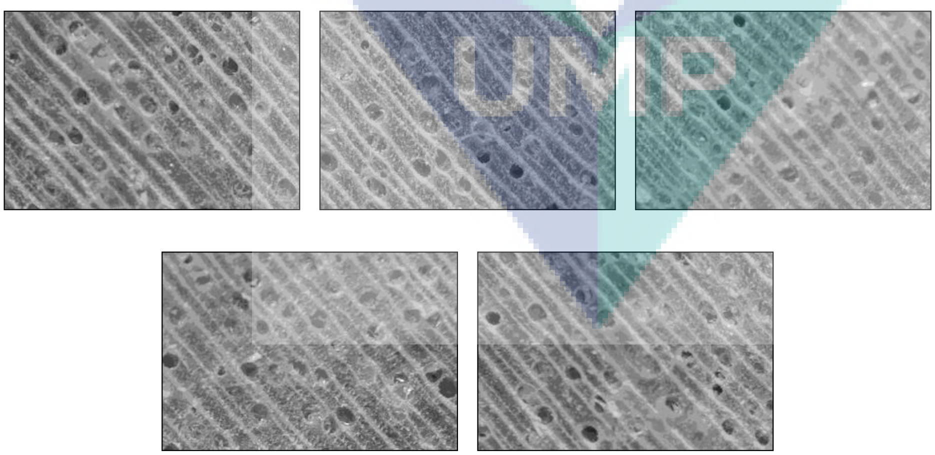
APPENDIX A5

UMP DATABASE: 5 IMAGES OF MERSAWA WOOD SPECIES

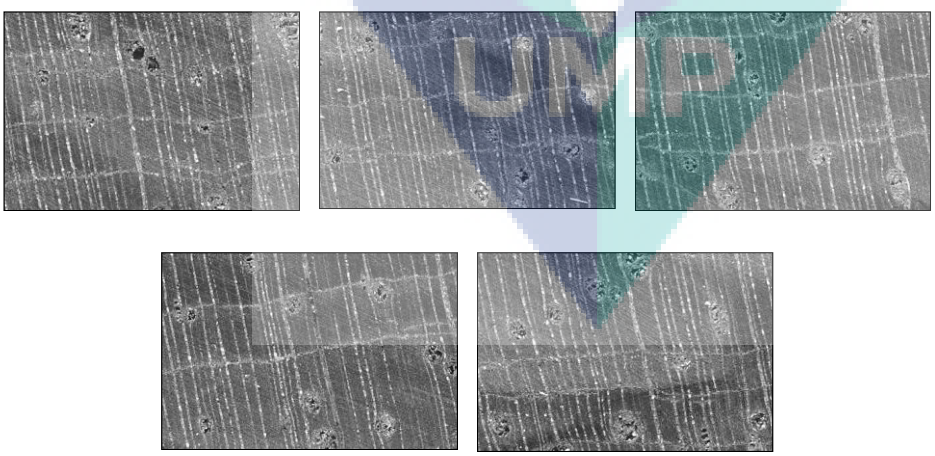
APPENDIX A6

UMP DATABASE: 5 IMAGES OF MERSAWA GAJAH WOOD SPECIES



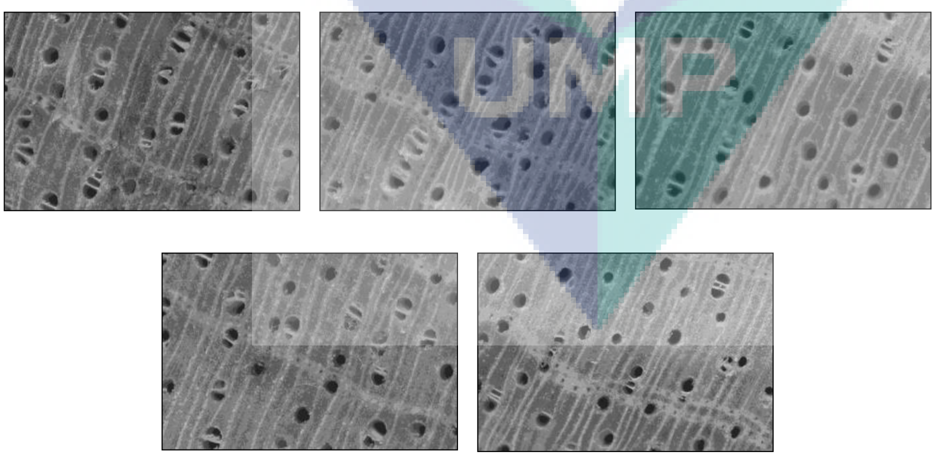
APPENDIX A7

UMP DATABASE: 5 IMAGES OF RENGAS WOOD SPECIES



APPENDIX A8

UMP DATABASE: 5 IMAGES OF SEPETIR WOOD SPECIES



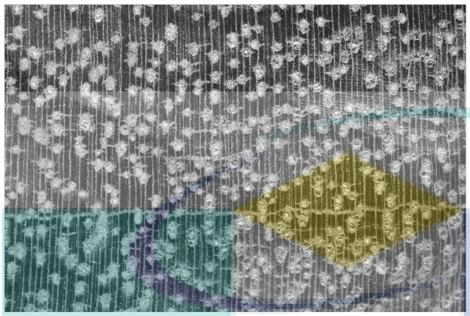
APPENDIX B

RESULT OF IMAGE ENHANCEMENT

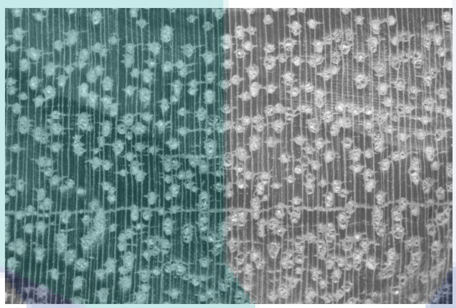
APPENDIX B1

WOOD SPECIE: BALAU

Original Image



Grayscale Image

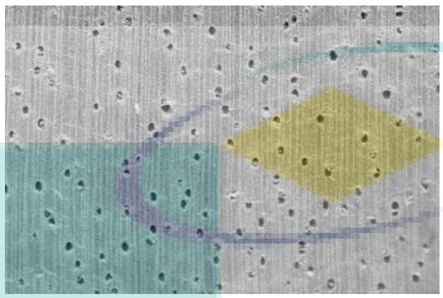


Average Filtering Image

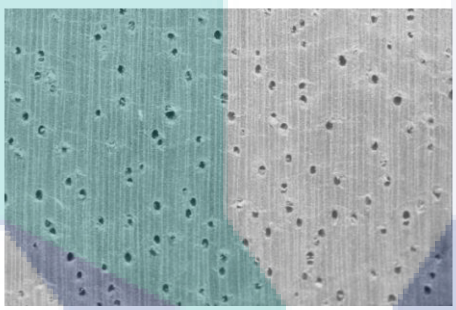
APPENDIX B2

WOOS SPECIES: RAMIN

Original Image



Grayscale Image

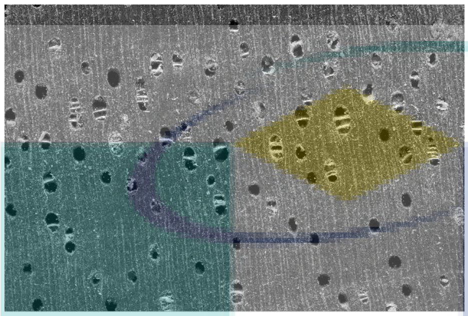


Average Filtering Image

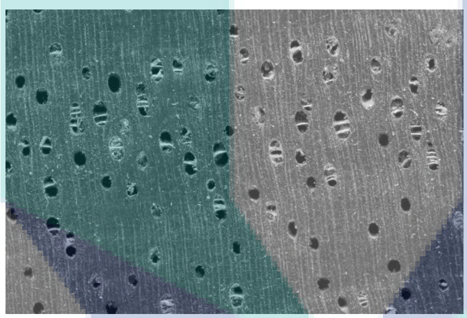
APPENDIX B3

WOOD SPECIE: DURIAN

Original Image



Grayscale Image

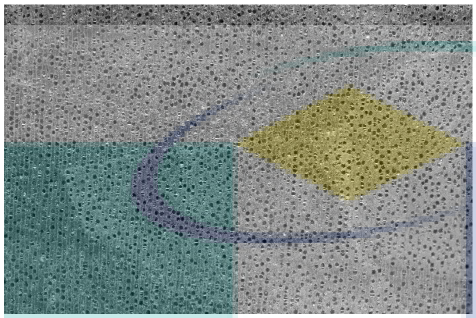


Average Filtering Image

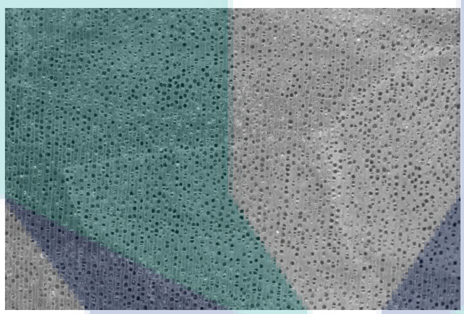
APPENDIX B4

WOOD SPECIE: MELUNAK

Original Image



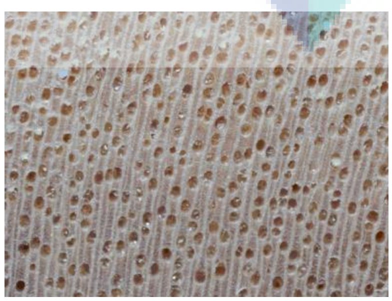
Grayscale Image

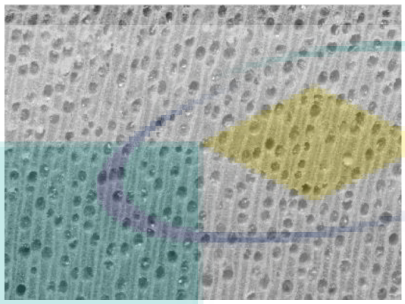


Average Filtering Image

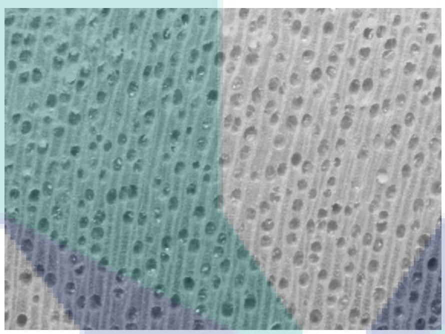
APPENDIX B5

WOOD SPECIE: MERSAWA





Original Image



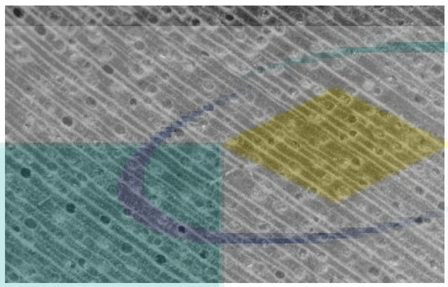
Grayscale Image

Average Filtering Image

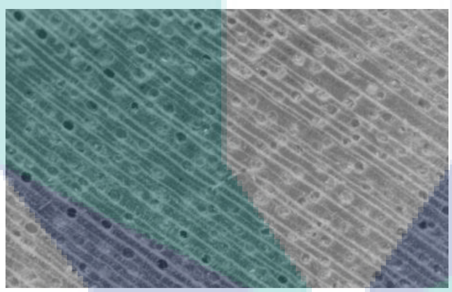
APPENDIX B6

WOOD SPECIE: MERSAWA GAJAH

Original Image



Grayscale Image

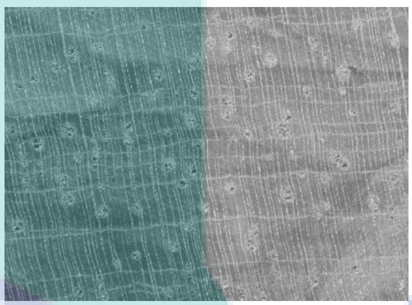


Average Filtering Image

APPENDIX B7

WOOD SPECIE: RENGAS

Original Image



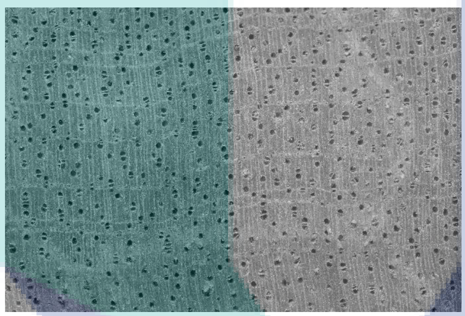
Grayscale Image

Average Filtering Image

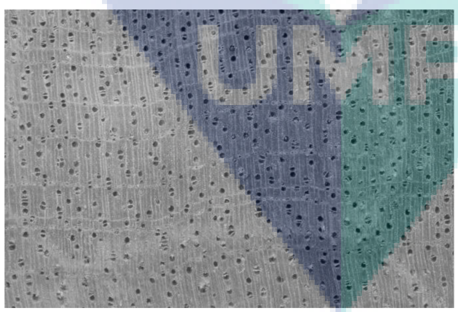
APPENDIX B8

WOOD SPECIE: SEPETIR

Original Image



Grayscale Image



Average Filtering Image

APPENDIX C

USER MANUAL

1. This system is started when the user open the MATLAB version 7.10.0 and above. The user will find the main interface of MATLAB. See Figure C – 1.
2. Then choose the project folder of the system.

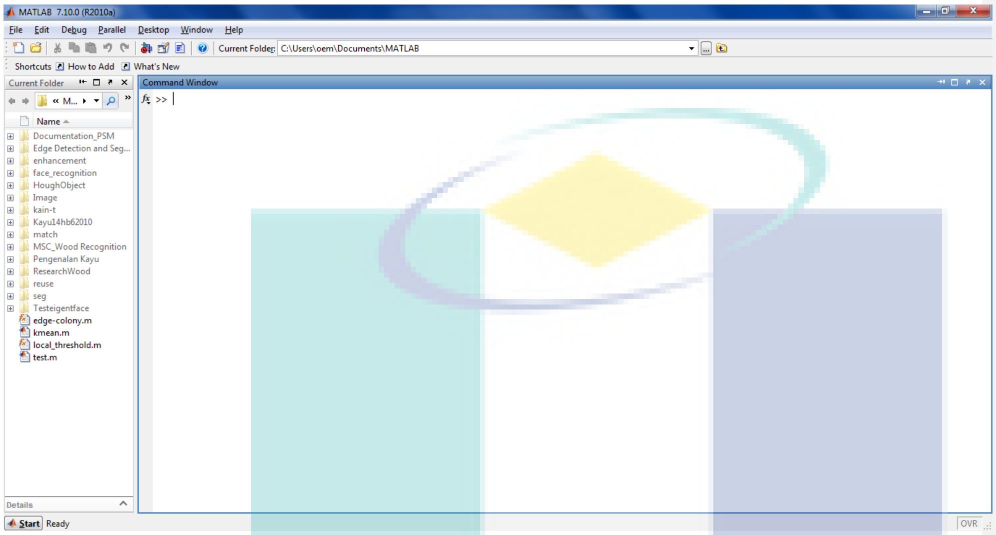


Figure C-1: Main interface of MATLAB 7.10.0

3. Open the SystemWoodRec folder and type “ Wood_Species_Recognition” in the command windows in the Figure see Figure C-2 or press button F5 and the program “Wood Species Recognition System “ has display in windows.

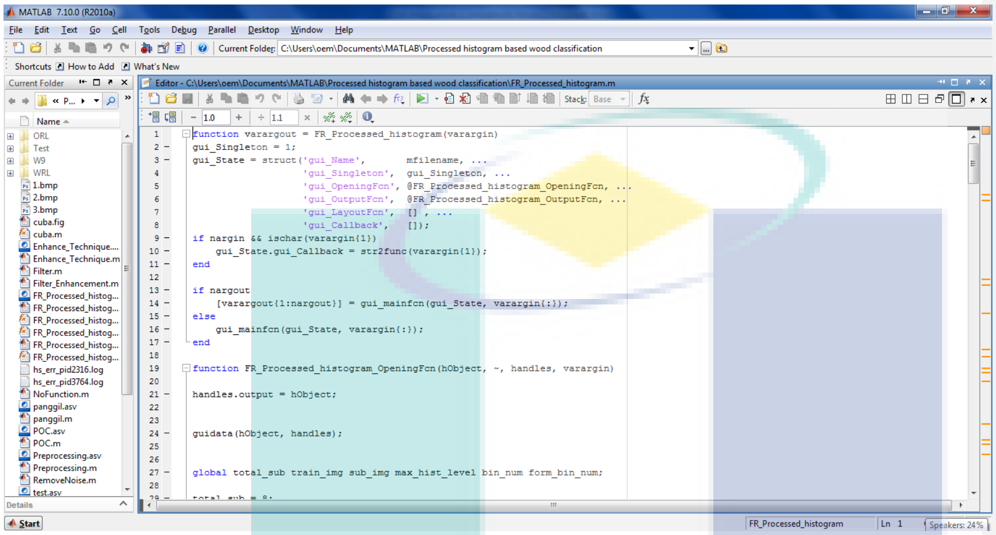


Figure C-2: Select the file path of the project

4. Main interface of the system has appeared. See Figure C-3.

Figure C-3: Main interface for system

5. When the main interface of the system appeared, the user need to click at the "TRAIN" button to train the images. See Figure C-4.

Figure C-4: Train image from database

6. Click "INPUT IMAGE" button to select wood species. See Figure C-5

Figure C-5: Select image from file

7. When the wood species image selected is appeared on the main interface of the system, the user needs to click at “TESTING” button to process the image that have been selected before.

8. The image will be processed. When all the process is done the result will be appeared on the main interface of the system. See Figure C-6.

Figure C-6: Interface for process and result