UNSUPERVISED IMAGE CLASSIFCATION USING ISODATA AND FUZZY C-MEANS

KEK SOO LING

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS TESIS

JUDUL: **UNSUPERVISED IMAGE CLASSIFICATION USING ISODATA AND FUZZY C-MEANS**

SESI PENGAJIAN: **SESI 2013/2014**

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

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

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

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

_____√_____TIDAK TERHAD

_____	_____
(TANDATANGAN PENULIS)	(TANDATANGAN PENYELIA)
Alamat tetap :3510, Taman Muhibbah, 73200 Gemencheh, NS	EN. OTHMAN BIN MOHD Nama Penyelia
Tarikh : _____	Tarikh: _____

CATATAN: * Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM)
** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.

UNSUPERVISED IMAGE CLASSIFICATION USING ISODATA AND FUZZY C-MEANS

KEK SOO LING

This report is submitted in partial fulfillment of the requirement for the Bachelor of Computer Science (Computer Networking)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA
2014

DECLARATION

I hereby declare that this project entitled

UNSUPERVISED IMAGE CLASSIFICATION
USING ISODATA AND FUZZY C-MEANS

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

STUDENT: _____ DATE: _____
(KEK SOO LING)

SUPERVISOR: _____ DATE: _____
(EN.OTHMAN BIN MOHD)

DEDICATION

Dear Parents

Thank you for your giving me the big support and encouragement. Your biggest support and care have helped me to achieve the final task in my university life.

Dear Lecturer, Supervisors and Evaluator

Thank you for your guidance, encouragement and knowledge.

ACKNOWLEDGEMENTS

Firstly, I feel lucky to have persevere, patience and good health throughout the duration of this final year project research. I feel very fortunate to have En Othman Bin Mohd as my research supervisor. Besides, I would like to express my high appreciation to my supervisor En. Othman Bin Mohd. The supervision and support that he gave truly help the progression and smoothness of my final year project. The co-operation is much indeed appreciated. For evaluator, Dr Azman Abu, thank you for guidance during the presentation on PSM and also for the evaluating and reading this report.

Moreover, deepest thanks and appreciation to my parents and family members for their encouragement and full of support for the report completion, from the beginning till the end.

Last but not least, thanks to all of my friends, they have been contributed by supporting my work and help myself during the final year project progress.

ABSTRACT

The research is focus on unsupervised of satellite image classification to classify an image into thematic image by using two techniques: Iterative Self-Organizing Data Technique Algorithm (ISODATA) and Fuzzy C-Mean (FCM). ISODATA is an extension of K-Means algorithm but ISODATA determines the number of clusters dynamically. ISODATA tries to find the best cluster centroids through the iterative approach until it meet some convergence criteria. Besides, ISODATA involves splitting and merging of the resulting clusters based on the user pre-specified thresholds. For splitting situation, when a cluster standard deviation above a pre-specified threshold, a cluster will be split into two while for merging situation, when the distance between the centroids is below another pre-specified threshold, two clusters are merged. FCM is the most popular fuzzy clustering technique which allows one data point to belong to two or more groups or clusters with different membership degrees between 0 and 1. The aim of FCM is to find the cluster centers that minimize a dissimilarity function. Thus the research is necessary to know the how the classification work and which of the techniques can produce the better output after classify the satellite image. Moreover, the research will compare the results generated between the two different techniques. The research is carried out using Matlab R2010a and at the end of this research, the result will show which techniques will produce the better output after doing the comparison of the final outputs.

ABSTRAK

Kajian ini fokus kepada teknik tanpa pengawasan ke atas imej satelite untuk membuat klasifikasi dengan mengunakan dua teknik iaitu ISODATA dan FCM. ISODATA merupakan lanjutan kepada KMeans algorithma tetapi ISODATA menentukan beberapa kumpulan secara dinamik. ISODATA akan mencari kluster sentroid yang terbaik melalui pendekatan leleran sehingga is mencapai penumpuan kriteria. Selain itu, ISODATA melibatkan pemisahan dan penggabungan keputusan kluster atas ambang yang dispesifikasikan oleh pengguna. Bagi situasi permisahan, kluster akan memisahkan kepada dua bahagian semasa sesuatu sisihan piawai kluster berada pada tahap yang atas ambang yang dispesifikasikan. Bagi situasi penggabungan, dua kluster akan bergabung semasa jarak antara sentroid kurang dari tahap yang ditetapkan. FCM merupakan teknik fuzzy yang popular kerana ia berkaitan dengan hubungan antara data dengan 0 and 1. Maka, objektif FCM adalah mencari kluster sentroid untuk mengurangkan ketidaksamaan antara algorithma. Kesimpulannya, kajian ini adalah untuk menjalankan pengklasifikasian dan membincangkan teknik yang menghasilkan keputusan yang baik dengan menggunakan Matlab 2010a.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TERMS	DESCRIPTIONS
FCM	Fuzzy c-Means
ISODATA	Iterative Self-Organizing Data Analysis Technique Algorithm
MSE	Mean Square Error
RP	Research Problem
RQ	Research Question
RO	Research Objective
TIFF	Tag Image File Format
uint 8	Unsigned 8-bit

CHAPTER I

INTRODUCTION

Nowadays, there are a large number of satellite images been generated and widely used in many field such as forestry, geology, military and so on. A satellite image commonly has multiple bands and different band representing different wavelengths from the ultraviolet. These ultraviolet wavelengths were through the visible and infrared portions of the electromagnetic spectrum. In other words, satellite images consist on layers and each layer contains special information that requires us to reveal (Arias et al., 2009). The data of the satellite image provides direct observation of the land cover at repetitive interval, thus allow class mapping and allow identifying the changes in land cover. The techniques that used to classify a satellite image can be categorized into two which are supervised and unsupervised classification techniques. Supervised classification techniques consist of Minimum Distance to Mean, Maximum Likelihood, Mahalanobis, Stepwise Linear, Suits' Maximum Relative and Back Propagation while unsupervised classification techniques consist of Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA), Fuzzy C Means, K-Means, Simple One-Pass Clustering, Minimum Distribution Angle, Self-Organization, and Adaptive Resonance. This project discusses on the use of two unsupervised classification techniques to identify the land cover using QuickBird satellite image. The main focus of this study is to classifying satellite image using ISODATA and Fuzzy C-Mean as well as to compare which technique produces better result or more accurate to represent the data in satellite image.

1.1 Project Background

Image processing is a method to extract the useful information from an image and by performing some algorithms or operations after converting the image into form of digital (Hossain, 2012). It was an assignment of signal in which the input image can be in form of photo or video frame and produce an output that full with characteristics that related to the input image (Hossain, 2012). Image processing can be divided into several parts which include image classification, image compression, image segmentation, image restoration and so on. However, this project will be covered on image classification.

Image classification is the process of extracting differentiated classes or themes from raw remotely sensed data. The main or objectives of the classification process is to categorize all pixels in an image with amount of digital data into several spectral classes, or themes with similar spectral attributes and then each spectral class is assigned to an information class (Caprioli et al., 2003).

Normally, multispectral data are used to perform the classification and, indeed, the spectral pattern present within the data for each pixel is used as the numerical basis for categorization (Lillesand and Kiefer, 1994). Categorization of image pixels is based on their digital number or gray values in one or more spectral bands.

As humans expert with satellite image based on visual perception that provide surroundings information but as sensors, it provide visible range of electromagnetic energy which are greatly limited sensitivity to humans eyes and it is able to form a lasting record of what human view, therefore a lot of methods been developed to increase the ability to see and record the physical properties of land cover (Tur et al., 2001).

Generally, there are two approaches taken to perform digital image classification which are supervised and unsupervised classification. For supervised image classification, it is required prior knowledge and familiar about the land cover and the spectral classes must be identified. The data from the pixels in an image are

used to train an algorithm of image classification. After trained, the algorithm can be applied to all of image and finally the image that is classified is gained (Kamaruzaman et al., 2009). Supervised classification was divided into several techniques which were Minimum Distance to Mean, Maximum Likelihood, Mahalanobis, Stepwise Linear, Suits' Maximum Relative and Back Propagation.

For unsupervised image classification, it does not require prior knowledge about the land cover and the image is automatically classified into spectral classes based on natural groupings found into the data (Caprioli et al., 2003). It was divided into several techniques which were Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA), K-Means, Fuzzy C Means, Simple One-Pass Clustering, Minimum Distribution Angle, Self-Organization, and Adaptive Resonance.

A pixel in satellite image might represent a mixture of class covers, variability of within-class or other complex surface cover patterns. The overlapping pixel cannot be correctly described by one class due to the image spatial resolution (pixel size). The mixed pixel (overlapping pixel) may influenced the accuracy of the classification or misclassification even though a good training area been selected (Mohd Othman et al., 2012).

According to (Ozesmi et al., 2002), the problem of overlapping pixels is dealing with fuzzy classification. In supervised classification, mixed pixels typically generate misclassification while the mixed pixels in unsupervised classification are dealing by using mixed labels because the information can be extracted without using the traditional per-pixel classifier (Mohd Othman et al., 2012).

However, this project was only focus on the unsupervised image classification using the two different techniques; ISODATA and Fuzzy C Means.

1.2 Problem Statements

Each pixel in an image reflects different spectral or represent important information that required us to reveal. Sometimes we have no prior knowledge about the land cover and it is not easy to recognize or identify the object or pixels inside the image that needed. In order to make the image easy to classified, recognized and identified without prior knowledge, the unsupervised image classification is applied. However, there are a lot of techniques under image classification and each produce different result. Therefore, it was required to find out which technique is more suitable. The Research Problem (RP) is summarized into Table 1.1 as below:

Table 1.1: Summary of Problem Statements

RP	Research Problem
RP1	Too many techniques can be applied to image classification.
RP2	Lack of understanding how unsupervised techniques process.
RP3	Lack of image understanding and recognition of an object or pixel which represent information in the image.

Based on the research problems in Table1.1, three research questions are conducted and the explanation for each of the Research Problems is explained as below:

RP1: Too many techniques can be applied to image classification.

This research problem is due to a lot of techniques that can be applied in image classification and may no prior knowledge about the techniques and have to make a decision to select suitable techniques to classify the satellite image

RP2: Lack of understanding how unsupervised techniques process.

This research problem is due to different techniques may produce different result or output. Therefore it is needed to understand and analyze how these unsupervised image classification techniques to classify the image.

RP3: Lack of image understanding and recognition of an object or pixel which represent information in the image.

This research problem is because of each pixel in an image can be a mixed pixel and unmixed pixel. In unmixed pixel, the data or information in an image can be easily be identified but in mixed pixels, there may contain two different pixels inside one pixel, therefore it is hardly to be classified.

1.3 Research Questions

Based on the problem statement that been stated in Table 1.1, the research question was identified. The research questions were summarizing and stated in Table 1.2.

Table 1.2: Summary of Research Questions

RP	RQ	Research Questions
RP1	RQ1	Which image classification techniques been selected to classify the satellite image?
RP2	RQ2	How unsupervised image classification techniques function?
RP3	RQ3	What is the meaning or information of the group of the pixels in an image that produced by the techniques represent?

From Table 1.2, the explanation for each of the Research Questions is constructed and described as follow:

RQ1: Which image classification techniques been selected to classify the satellite image?

This research question is stated in order to understand which classification techniques are suitable for this project. Without a prior knowledge about the land cover, unsupervised image classification is more suitable in this project

RQ2: How unsupervised image classification techniques function?

This research question is stated by considering the understanding of the procedure of how the unsupervised image classification techniques function. Each unsupervised technique performed in different ways to classify an image.

RQ3: What is the meaning or information of the group of the pixels in an image that produced by the techniques represent?

This research question is stated in order to understand and recognize the groups of pixels that been classified and the pixels that bring out the meaningful information to the image.

1.4 Objective

The appropriate research objectives are constructed based on the research problems and research questions objectives that were mentioned before. The summary of the research objectives were shown in Table 1.3.

Table 1.3: Summary of Research Objectives

RP	RQ	RO	Research Objectives
RP1	RQ1	RO1	To select different techniques of unsupervised image classification to separate land cover from satellite image.
RP2	RQ2	RO2	To analyze unsupervised classification techniques for satellite image processing.
RP3	RQ3	RO3	To separate land cover of satellite image using different unsupervised classification techniques

From the Table 1.3, the explanation of each of the research objectives are constructed as follow:

RO1: To select different techniques of unsupervised image classification to separate land cover from satellite image.

In order to classify the pixels of an image into meaningful data, first need to identify the image classification techniques. Since do not have the prior knowledge of land cover, unsupervised classification techniques are more suitable to be selected. The unsupervised classification techniques that been chosen were ISODATA and Fuzzy C-Means.

RO2: To analyze unsupervised classification techniques for satellite image processing.

Each unsupervised techniques have their own procedure and function, therefore need to analyze the procedure of each technique to classify the pixels into meaningful class.

RO3: To separate land cover of satellite image using different unsupervised classification techniques

Based on the analysis, each data in an image will be separate into different clusters and the output will be produce.

1.5 Scope

The scope of this project focusing on the ISODATA and Fuzzy C-Means techniques which is used to compare and find out the most suitable unsupervised image classification which produce more accurate output. The QuickBird satellite image that used in this project was using uint8(unsigned integer 8) with the size of 225x277x3 pixels from Erdas Imagine. The satellite image was taken on 17 March 2010 of Kuala Linggi Mangroves Forest, Alor Gajah Melaka. The software used to construct the studied was Matlab R2010a.

1.6 Project Significant

The main purpose of this project was to automatically classify the image into spectral classes based on identification of natural groupings found in the data. The procedure of classifying the satellite image using ISODATA and Fuzzy C Mean were shown in this project and the thematic classes been produced. At the end of this project, the resulted of image classification using these two techniques were analyzed and compared. Finally, each percentage of the generated clusters will be calculated.

1.7 Expected Result

By the end of this project, the expected results must achieve the goals of this project:

i. The procedure of each unsupervised image classification techniques that performed to classify the satellite images is shown.
ii. The objects or pixel in an image can be distinguished and identified easily.
iii. The thematic image is identified and shown.

1.8 Report Organization

In this project, the report organization included six main chapters. Each of the main chapters will describe the main task or procedure in order to perform this project.

In Chapter I: Introduction, eight main topics which included the project's background, problem statements, research questions, project objective, project scope, expected result, and report organization were discussed in this chapter. The ideas of this project were produced through the research question and problem statement that stated and finally the objectives of the research are generated. Besides that, project background is an important part to get an idea on the flows of the project. The project

scope was focus on the domain of the project which was ISODATA and Fuzzy C Means. Apart from that, the expected result is the part that stated the goals of the objective and must be achieved at the end of the project. Finally was the report organization which was carried out to assure the report was organizing in a systematic manner.

In Chapter II: Literature Review, the related work, analysis of current problem and justification, and proposed solution for the project were discussed and described in this chapter. The related work was the research on the basic concepts of the project domain. After that, current problems in the current situation of image classification were analyzed and discussed. In order to overcome those problems, the proposed solutions were proposed.

In Chapter III: Methodology, the topics included project methodology, project schedules and milestones were discussed. The project methodology shows the overall of the project flows and the main five steps flows were discussed in this chapter. For the project schedules and milestone, it was a topic that shows the person that included in this project and how to organize the time and activities in order to perform in a timely manner.

In Chapter IV: Implementation, the two main topics were discussed: project requirement and the algorithms of the unsupervised classification techniques. In project requirement, it was divided into two sub topics which were software requirement and hardware requirement while for algorithms of the unsupervised classification techniques, it was discussed the algorithms and the different processing flows of the classification techniques and the steps to implement in this project were discussed in this chapter.

In the Chapter V: Testing and Analysis, test plan and test description are the main topics for this chapter. The test plan was taken to find out the person that responsible to test this project. The test description is the step by step testing that used for the two classification techniques.