Republic of Iraq

Ministry of Higher Education

and Scientific Research

University of Diyala

College of Science

Satellite Image Classification Using Data Mining Technique

A Research

Submitted to College of Science/ University of Diyala in a Partial Fulfillment of the requirements for the Degree of Master in computer Science

By

Ghaidaa Waleed Naji

B.Sc. in Computer Science / College of Science / Diyala University (2013)

Supervised By

Assist. Prof. Dr. Jamal Mustafa Al-Tuwaijari

2018 A.D. **1439 A.H.**

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

وَمِنْ آيَاتِهِ خَلْقُ السَّمَاوَاتِ وَالْأَرْضِ وَمَا بَثَّ فِيهِمَا
مِنْ دَابَّةٍ وَهُوَ عَلَى جَمْعِهِمْ إِذَا يَشَاءُ قَدِيرٌ ۝

صَدَقَ اللهُ العَظِيم

سورة الشورى الآية (29)

Acknowledgement

Firstly, all my prayers go to (Allah), the Almighty, for the successive blessings, divine providence, and my success in this research.

Special Thank and gratitude should be extended to my supervisor ***Assist. Prof. Dr. Jamal Mustafa Al-Tuwaijari*** *for his tireless patience, useful comments, and excellent advice and support.*

I am grateful to the staff of computer science, who always takes the necessary steps to ease the difficulties that stand in the way of research.

Last and not least, thanks a lot go to my ***family, my friends, and anyone*** *who helped me in one way or another.*

Dedication

To my Father, who embraced me with his love and guidance,

To my Mother, she patiently supported me to see my success

To my sisters

To my Brother,

I produce this work with all my love.....

Ghaidaa

Linguistic Certification

I certify that this research entitled **"Satellite Image Classification Using Data Mining Technique "** was prepared by **Ghaidaa Waleed Naji** and was reviewed linguistically. Its language was amended to meet the style of English language.

Signature:
Name: Assist.Prof.Dr. Mohammad Naji Hussain
Date: / / 2018 11/12

Supervisor's Certification

I certify that this research entitled "**Satellite Image Classification Using Data Mining Technique** " was prepared by **Ghaidaa Waleed Naji** under my supervision at the Department of Computer Science University of Diyala as a partial fulfilment of requirements for the Degree of **Master of Science in Computer Science.**

Signature:

Name: Assist. Prof. Dr. Jamal Mustafa Al-Tuwaijari

Date: ١/ ١٧ 2018

Approved by the computer sciences department, college of science, university of Diyala.

Signature:

Name: Assist. Prof. Dr. Taha Mohammed Hasan

Date: ١ / ١١/2018 (Head of the Department)

Examining Committee Certification

We certify that we have read this research entitled "***Satellite Image Classification Using Data Mining Technique***", and as an examining committee, examined the student ***Ghaidaa Waleed Naji*** in its contents and in our opinion, it has adequate fulfill the requirements for the Degree of Master of Science in Computer Science department.

(Chairman)

Signature:

Name: Dr. Dhahair A. Abdulah Salman

Title: Professor

Date: 1 / 11/2018

(Member)

Signature:

Name: Dr. Abbas Fadhil Mohammed Ali

Title: Assistant Professor

Date: 1 / 11 /2018

(Member)

Signature:

Name: Naji M. Sahib

Title: Assistant Professor

Date: 1 / 11/2018

(Member /Supervisor)

Signature:

Name: Dr. Jamal Mustafa Al-Tuwaijari

Title: Assistant Professor

Date: 1 / 11/2018

Approved by the Dean of college of science, university of Diyala.

(The Dean)

Signature:

Name: Dr. Tahseen Hussein Mubarak

Title: Professor

Date: 29/ 11/2018

Abstract

Satellite images constituted of photographs of earth produced by artificial satellites. Satellite imagery is involved in many areas where it has many applications in meteorology, agriculture, geology, regional planning, biodiversity conservation, forestry, intelligence, warfare, and education. classification is a significant technique applied in remote sensing for the analysis and pattern recognition of satellite data, which assist the automated interpretation of a massive amount of information.

Today, there are many kinds of classification algorithms. Data mining techniques is an important field in computer science and used as computational process of defining patterns in large data. In this thesis, two types of data mining techniques are used which is supervised classification and clustering. Satellite images with different resolution are adopted in this thesis which is satellites images with medium resolution and high resolution. In high resolution images takes as scenes. The supervised classification used to classify high-resolution satellite images as scene classification by using two techniques support vector machines and K-Nearest Neighbor's with 20 class. The accuracy computed for each algorithm and the test results show that the proposed classification methods obtain very favorable performance, where 20 classes were classified and the highest classification accuracy for one of the classes was 95% for SVM algorithm.

Clustering used to classify medium resolution images (Landsat-8). Different types of algorithms have been applied for clustering, which are k-means clustering and fuzzy c-means with two approaches pixel based clustering and block based clustering. The pixel based approach depend on pixel intensity for clustering while in block based clustering color features and texture features are extracted. In texture features gray level co-occurrence matrix (GLCM) is used. Finally, the obtained results are used for comparison between the two algorithms which showed that the performance of k-means clustering algorithm

was better than the performance of fuzzy c-means clustering algorithm, the best accuracy obtained was 92.4194 for image C using k-means algorithm while in fuzzy c-means algorithm the best accuracy obtained was 87.1384 for image D.

Table of Contents

List of Abbreviations

Abbreviation	Description
SPOT	Satellite Pour l'Observation de la Terre
RBV	Return Beam Vidicon
MSS	Multi Spectral Scanner
TM	Thematic Mapper
ETM	Enhanced Thematic Mapper
OLI	Operational Land Imager
TIRS	Thermal Infrared Sensor
GIS	Geographic Information Systems
RGB	Red,Green,Blue
FE	Feature Extraction
GLCM	Gray Level Co_occurrence Matrix
σ	Standard Deviation
μ	Mean
LBP	Local Binary Patterns
MI	Mutual Information
SVM	Support Vector Machine
KNN	K-Nearest Neighbor

List of Figures

List of Tables

Chapter One

General Introduction

Chapter One

General Introduction

1.1 Introduction

Today satellite images play an important role in giving the geographic information needed for different purposes. Satellite and remote sensing technologies gather images at uniform intervals. The technology is growing rapidly and the amount of data is growing at a great rate where the amount of data received at data centers is massive and growing exponentially. To extract useful information and interpret it from large quantities of satellite images, we need a powerful and sophisticated mechanisms. Classification of satellite image consider robust mechanism to extract data from a massive amount of satellite images. The classification of satellite images includes the compilation of pixels into importance classes. The process of classifying satellite images requires from analyst to make a lot of decisions and choices, which makes the classification of satellite images somewhat complex [1]. A lot of geographic information is present in remote sensing images. Geographic information can be important in many areas such as business, science, research, and governments, it can also be used in city planning and natural resource analysis in certain areas [2]. Image classification include of database that contains predefined patterns and compares those patterns to discover object to classify it into suitable group, appropriate algorithms are used to determine target images from an image database and introduced image correctly and accurately. Procedures of image classification comprise of preprocessing, feature extraction, feature selection, and classification. At present day data mining techniques are used to process computer vision problems such as image classification, object recognition, and etc. The fundamental objective of data mining technique is to improve the process of getting patterns which should be

efficient, scalable and can detect the important patterns which can be exploit in different ways [3].

1.2 Remote Sensing

Remote sensing is the science of obtaining information about the surface of the Earth without real contact with it. This is accomplished by sensing and recording the energy emitted, processing and analyzing this energy and applying the information obtained.

Remote sensing includes, in a broader sense, satellites, air monitoring and spacecraft for the surfaces and planetary structures in our solar strands where earth is the most common target of study .There are many purposes related to remote sensing of the Earth, including the work and updating of maps, weather forecasting and military purposes [4].

In the beginning, aerial photography was used, and then remote sensing was seen as an important tool for viewing, analyzing, distinguishing, and making decisions about our environment .In the past decades remote sensing has advanced in three important aspects: 1) from military use to various environmental applications associated with ocean, land and land issues 2) from photographic imaging systems to the more sophisticated use of sensors 3) and finally is the evolution of aircraft to the use of satellite [5].

In remote sensing, the process involves an interaction between the falling radiation and the target [6]. This process is summarized in the following seven steps and each step shows a different process from the other, as show in figure (1.1) and will explain in the following seven elements:

A. **Illumination**: The first requirement for remote sensing is to have a source of energy that provides electromagnetic energy to the intended target. This process is called the energy source.

B. **Radiation and the Atmosphere:** This component represents the energy transmitted from the source to the target as this energy will be in contact with the atmosphere through which it passes. This interaction occurs for the second time when the energy moves from the target to the sensor.

C. **Interplay with the Target:** As soon as energy takes its way to the target passing through the atmosphere, it interacts with the target depending on the characteristics of both the target and the radiation.

D. **Record Energy using the Sensor:** As mentioned above, the target is energy-generating, since this energy from the target must be invested in this situation. We need a device that is not in real contact with the target called a sensor to collect and register electromagnetic radiation and to benefit from it.

E. **Transmit, Receipt and Handling Data:** After the energy has been received by the sensor and recorded, this recorded energy must be sent by the sensor to the receiving station electronically transmitted. The received data is processed in digital or printed form.

F. **Interpret and Analysis Data:** To extract information about the target that triggers energy, the processed image is interpreted and analyzed electronically, digitally or visually.

G. **Application:** The last step in the process of remote sensing is the application, that is, we apply the information obtained from the satellite about the target in order to provide a better understanding of it, as well as exploring new information which help to solve a certain problem [6].

Figure 1.1 illustrate all these steps.

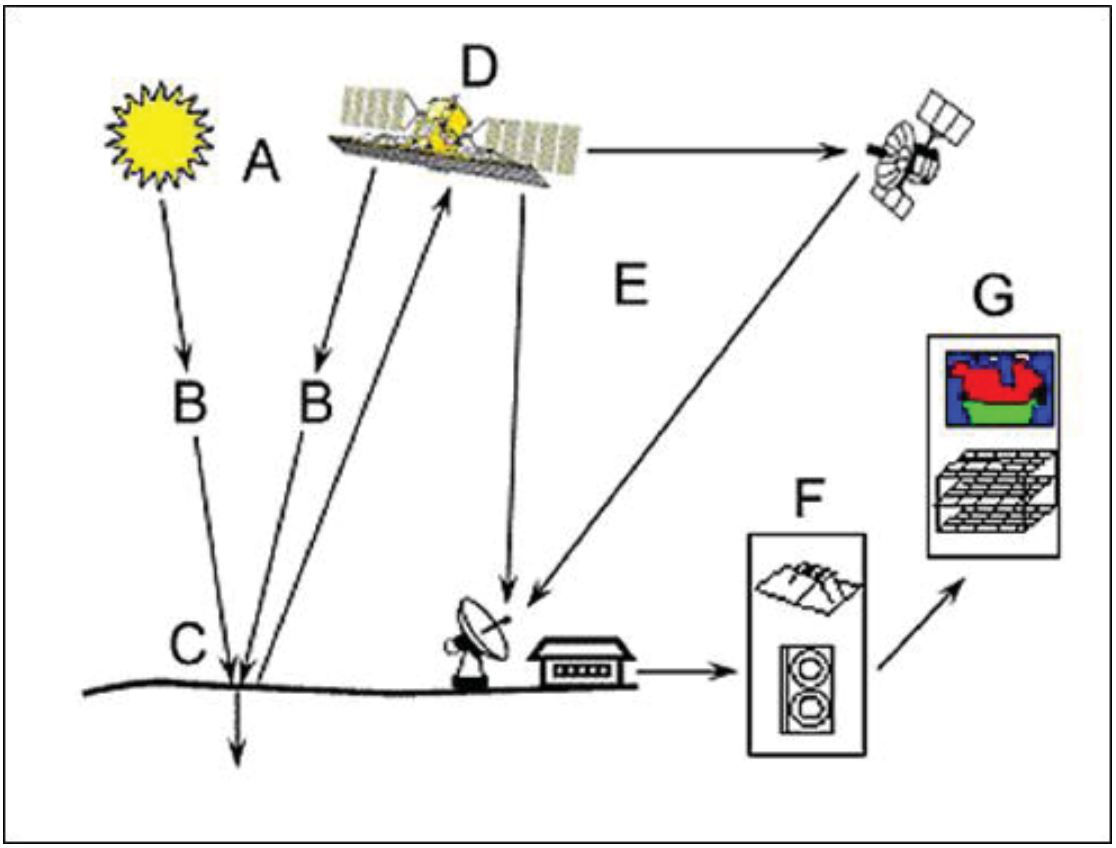

Figure 1.1: Steps of remote sensing process [6].

1.3 Satellite Imagery

At present, many companies and governments are operating satellite remote sensing systems, made or specially designed to monitor the earth's surface, to collect information related to issues such as water bodies, land, minerals, cities, etc. The satellite is distinguished from the aerial platforms with many advantages, as it can provide a comprehensive vision, which means the observation of large areas in one large image and get more details and more, as well as repeated methodological coverage. There are different type of satellite imagery with different resolution such as Landsat, Satellite Pour l'Observation de la Terre (SPOT), IRS, Ikonos , Quickbird , Worldview , ALOS , Geoeye ,[7] figure 1.2 illustrate some types of these satellite .

A. Landsat 8

B. Worldview3

C. IKONOS

D. Quikbird

Figure 1.2: Some types of Satellite Imagery [8]

1.3.1 Landsat

It considers one of the most famous remote sensing satellite families, launched from the United States and developed over time. In 1972, the first satellite of this family was launched, called Landsat 1 with two sensors and the final satellite of this family is Landsat 8 which launched in 2013 with 11 bands. Table 1.1 show the history of this family [9].

Table 1.1: Landsat History

Satellite	Lunch date	resolution	(days)
Landsat 1	7/23/1972	RBV 80; MSS 80	18
Landsat 2	1/22/1975	RBV 80; MSS 80	18
Landsat 3	3/5/1978	RBV 30; MSS 80	18
Landsat 4	7/16/1982	TM 30; MSS 80	16

Landsat 5	3/1/1984	TM 30; MSS 80	16
Landsat 6	10/5/1993	Pan 15; ETM 30	16
Landsat 7	4/15/1999	Pan 15; ETM +30	16
Landsat 8	2/11/2013	OLI , TIRS 30	16

1.4 Literature Survey and Related Work

Various techniques are proposed for the classification of satellite images. This section provides a set of previous techniques that dealt with the classification of satellite images:

S.V. S Prasad, T. Satya Savitri, I.V. Murali Krishna, 2011 [10]: In this research, an effective classification method for satellite images was proposed, with the help of both the SVM and clustering algorithm. The proposed work consists of four stages: pre-processing, segmentation, SVM training and finally classification by using SVM. In the pre-processing phase, Gaussian filtering used, in segmentation stage, fuzzy incorporated hierarchical clustering technique is applied, and then the output of this algorithm becomes an input to the SVM, where it is classified into cover used and land used according to the training data.

Guofeng Sheng, Wen Yang, Tao Xu a & Hong Sun, 2012 [11]: This research deals with the method of classification of high-resolution satellite images scenes. A variety of methods have been used to extract features which include three techniques: color histograms, LTP-HF (Local Ternary Patterns histogram Fourier) and SIFT (scale invariant feature transform). These features were combined to obtain a high classification accuracy. The first two stages were combined using the linear SVM classifier, which was first used to generate the probability of images using the three methods mentioned. The resulting probability images in the second stage are used for classification. Through the proposed classification methods, a great performance was achieved for the classification.

Shivali A. Kar, Vishakha V. Kelkar, 2013 [12]: In this research, satellite images were classified by using several techniques of supervised and unsupervised classification. A variety of features were extracted, such as mean, variance and standard deviation. The supervised classification was used for backpropagation and radial basis function, and the unsupervised classification used Self-organizing map.

Bharathi S, P Deepa Shenoy, Venugopal K R, L M Patnaik , 2013 [13]: In this work different resolution of images are considered. The complex features are extracted by using the combination of color and texture. K-means cluster is used to cluster the features and labeled the dataset. SVM, Bayes and neural network classifiers are applied. The classification is simple one using only five categories; water class, empty land class, farmland class, buildings class, trees class. The validation is done by field survey. The proposed method gives more than 90% of accuracy when it compared with the field survey.

Manali Jain, Amit Sinha, 2015 [14]: Introduce a way to classify satellite images through Gabor Filter using SVM. The focus of this work was on five class of satellite images which is desert, forests, mountains, residential areas and agriculture. A collection of features was used in this research which is a feature of color, texture, and shape. These features were extracted through the use of Gabor Filter. As for the classification algorithm used, it is an SVM algorithm with the use of RBF. Accuracy obtained was 98.5 % for satellite images.

Balamurugan G., K. B. Jayarraman,2016 [15]: In this research, the classification of land cover area from the test satellite Image employs first order statistics and Gray Level Co-occurrence Matrix with wavelet transformation. The classification progress has been achieved by supervised and unsupervised classification technique. The suggested scheme of wavelet transformation provides an effective mapping of land cover analysis for remote sensing.