UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Facultad de Filosofía y Letras
Posgrado de Geografía

COMPARACIÓN DE DISTINTOS MÉTODOS DE CLASIFICACIÓN DIGITAL DE IMÁGENES DE SATÉLITE

TESIS

Que para obtener el grado de:

Doctor en Geografía

Presenta

Y A N G A O

Director de Tesis
Dr. Jean François Mas Caussel

Revisores
Dr. José Luis Palacio Prieto
Dr. Gerardo Bocco Verdinelli
Dr. Raúl Aguirre Gómez
Dr. Diego Fabián Lozano García

Agosto 2008

IMAGE SEGMENTATION AND OBJECT BASED IMAGE ANALYSIS USING REMOTE SENSING IMAGES

Yan Gao

AGRADECIMIENTOS:

Esta tesis es el resultado de cuatro años de trabajo, en los que me han acompañado y apoyado muchas personas e institutos. Es un gran placer el que tenga ahora la oportunidad para expresar mi gratitud a todos ellos. En primer lugar, quisiera agradecer sinceramente al Fondo Sectorial CONACYT-CONAFOR por su apoyo para el proyecto "Evaluación del sensor MODIS para el monitoreo anual de la vegetación forestal de México" (clave 14741), y al Fondo Sectorial de Investigación Ambiental SEMARNAT-CONACyT (clave 2002-C01-0075), a ambas gracias por el apoyo financiero durante el periodo de mi doctorado.

También quiero agradecer profundamente doctor Jean François Mas, mi tutor principal, por su excelente supervisión, orientación y confianza. Siempre conté con su apoyo para el desarrollo de este trabajo, con detalles técnicos o respuestas a preguntas teóricas. Su opinión crítica me ha ayudado a mantenerme enfocada a lo largo de los años que abarcó mi investigación.

Estoy muy agradecida de la ayuda que me brindó mi segundo tutor, el doctor José Luis Palacio Prieto, por la revisión de mi tesis. El siempre ha estado allí cuando he necesitado su ayuda y estoy agradecida de su apoyo y confianza.

Quisiera también dar mi muy sincero agradecimiento al doctor Gerardo Bocco Verdinelli por aceptar ser mi tercer tutor y por esas muy interesantes preguntas que me formuló en mi examen de candidatura, así como por la revisión de mi tesis. Gracias por su apoyo y su confianza.

Estoy agradecida por la ayuda del maestro en ciencias José Antonio Navarrete Pacheco por su apoyo en la verificación de la exactitud de la clasificación de imágenes, así como en la elaboración de algunas de las figuras de esta tesis.

Me gustaría dar las gracias a la Profesora Doctora Imgard Niemeyer y Prashanth Marpu del Instituto de Minas, Tecnología y Geodesia de la Universidad de Minería y Tecnología de Freiberg, Alemania. Doctora Imgard Niemeyer supervisó mi estancia de investigación durante los cuatro meses que estuve en Freiberg, Alemania; y Prashanth Marpu, estudiante de doctorado, contribuyó en el desarrollo del capítulo 5.

Al doctor Norman Kerle del ITC, los Países Bajos, Gracias por su orientación para la realización del capítulo 3. Es un gran benefició para mí su orientación y opinión durante este trabajo.

Mi sincero agradecimiento al Dr. Raúl Aguirre Gómez y Dr. Fabián Lozano

García por sus valiosas opiniones para mejorar el manuscrito de esta tesis al aceptar fungir como sinodales de la misma.

Un agradecimiento a la Dra. Claudia Kuenzer por su apoyo a mi trabajo.

Me gustaría dar las gracias a mi familia, mis amigos y compañeros por su amor, amistad y compañía.

Un agradecimiento especial a Antonio Navarrete, sin su apoyo, esta investigación y la realización de la tesis no habría sido posible.

Y, por último, quisiera dedicar esta tesis a la memoria de México, espero haber brindado con mi trabajo una importante contribución a la ciencia de este país y para aquellos que son parte de él.

Yan Gao
Morelia, México, 2008

ACKNOWLEDGEMENTS

This thesis is the result of four years of work whereby I have been accompanied and supported by many people and institutes. It is a great pleasure that I have now the opportunity to express my gratitude for all of them. In the first place, I would like to thank Fondo Sectorial CONACYT-CONAFOR through the project "Evaluación del sensor MODIS para el monitoreo anual de la vegetación forestal de México" (clave 14741), thank Fondo Sectorial de Investigación Ambiental SEMARNAT-CONACYT (clave 2002-C01-0075), for supplying financial support throughout the PhD period.

I would like to express my sincere thanks to Dr. Jean François Mas, my first supervisor, for the excellent and challenging supervision, guidance through my research, fruitful discussions and his trust. His critical but motivating opinion helped me to stay focused throughout the PhD period.

I am grateful of the help from my second supervisor Dr. José Luis Palacio Prieto for the revision of my thesis. He would always be there when I needed his help and I am grateful of his support and trust.

I would like to give my sincere thanks to Dr. Gerardo Bocco Verdinelli, for accepting to be my third supervisor, for the revision of my thesis, and for those challenging questions in my PhD candidate exam. He has supported me in many ways and showed me his trust in my research ability.

I am grateful of the help from MSc. José Antonio Navarrete Pacheco for his support with the accuracy assessment of some of the classification results in this work and his help with elaborating the figures in some of the chapters (mentioned in the chapters).

I would like to thank Professor Imgard Niemeyer and then Phd student Prashanth Marpu from the Institute of Mine-Surveying and Geodesy at Freiberg University of Mining and Technology, Germany. Professor Imgard Niemeyer supervised my research during the four months of research stay in Freiberg, and Prashanth contributed to the development of chapter five.

I would like to thank Dr. Norman Kerle, from ITC, the Netherlands, for his guidance through the completion of a previous version of chapter three. I am greatly benefited from his challenging guidance, critical but motivating opinion throughout the working period.

Sincere thanks go to Dr. Raul Aguirre Gómez and Dr. Fabián Lozano García for accepting to be my exterior supervisors and thanks for their valuable opinions

to improve the manuscript of this thesis.

Thanks also go to Dr. Claudia Kuenzer for her guide, support and encouragement of my work.

I would like to thank my family, my friends and colleges for their love, friendship, and accompany.

Special thanks go to Antonio Navarrete, without his support, this research and the completion of the thesis would not have been possible.

And finally, I would like to dedicate this thesis to the memory of Mexico, and to those who are part of it.

Yan Gao
Morelia, México, 2008

RESUMEN

Esta investigación se propuso analizar el efecto de la segmentación de imágenes en la exactitud de la clasificación por objetos, y demuestra con varias aplicaciones como la segmentación de imágenes y la clasificación basada en objetos obtienen mejores resultados que los métodos basados en clasificación por pixeles, aplicando diferentes métodos de clasificación por objetos sobre imágenes de diferentes resoluciones. El nuevo algoritmo SEaTH, que calcula semi-automáticamente los parámetros y umbrales óptimos para la diferenciación de clases de cobertura, fue probado en la creación de cartografía de cobertura del suelo, y el Índice de Vegetación Mejorado (IVM) del sensor MODIS fue utilizado para examinar su contribución a la clasificación de imágenes basada en objetos.

El concepto principal de la clasificación de imágenes por objetos radica en que la información necesaria para interpretar una imagen no está contenida en un solo píxel, sino en los objetos presentes en la imagen, y en las relaciones entre los objetos. Dos motivos principales impulsan el desarrollo del análisis de imágenes por objetos: 1) Imágenes de alta resolución espacial ya están disponibles y las herramientas de computación están constantemente mejorando; 2) El análisis de la imagen basado en píxeles es limitado.

La segmentación de imágenes delinea los objetos de la imagen y es el paso inicial en el análisis de imágenes basado en objetos. Sólo hasta hace poco tiempo la segmentación de imágenes se empezó a realizar con un alto nivel de precisión y rapidez de ejecución, lo cual ha impactado en el desarrollo del análisis de imágenes basado en objetos. El análisis de la imagen por objetos está mucho más cerca de la percepción/visión humana y los resultados de la clasificación muestran una alta exactitud, y permiten leyendas más detalladas.

El proceso de segmentación divide la imagen en objetos homogéneos espectral y espacialmente. La evaluación de los resultados de varias segmentaciones de la imagen antes de realizar la clasificación de la misma asegura que el mejor resultado de segmentación se utiliza en la clasificación. Al realizar y evaluar

múltiples segmentaciones de una imagen Landsat_7 ETM+ con diferentes parámetros, el resultado de segmentación óptimo puede ser identificado. Los resultados de la segmentation se evalúan con una función objetiva y la segmentación con óptimo resultado es señalada. Mediante la clasificación de la imagen con el resultado de diferentes segmentaciones y la evaluación de su exactitud, se comprueba que la mejor segmentación, en términos de la función objetiva, también dio lugar a la clasificación con la más alta exactitud, y la distribución de los valores de exactitud de las clasificaciones presentan similares tendencia con respecto a los valores de la función objetiva de las segmentaciones.

El desempeño de la clasificación de las imágenes basado en píxeles y basado en objetos se comparó usando imágenes de satélite con diferentes resoluciones espaciales: 10m, 30m, 100m y 250m. Los resultados mostraron que con el misma información de entrenamiento y referencia, el análisis de imágenes basado en objetos ha obtenido mayor exactitud que los métodos basados en píxeles con las imágenes de mayor resolución espacial (10m y 30m), mientras que, con la reducción de resolución espacial, la clasificación de imágenes basada en objetos no mostraron mayor exactitud. Este experimento parece sugerir que el analysis de imágenes basado en objetos tiene ventajas sobre el basado en píxeles. Si bien en el aspecto de la exactitud de la clasificación, la ventaja sólo es válida para imágenes con altas resoluciones espaciales.

Los objetos de la imagen pueden ser diferenciados por una variedad de características: espectral, espacial, de textura, y del contexto. Un análisis completo de dichas características es esencial para trabajar con los objetos de la imagen, y el proceso de prueba y error de búsqueda de las características adecuadas podría dificultar significativamente la utilización de las bondades del análisis de imagenes basado en objetos. El algoritmo SEaTH o “SEperability & THreshold”, es capaz de evaluar estadísticamente un número determinado de características de los objetos para cualquier número de clases de interés. Este experimento con SEaTH muestra que el análisis de imágenes basado en objetos realizado con las características y umbrales identificados por el algoritmo SEaTH produjo buenos resultados de clasificación. Con SEaTH, el tiempo de prueba y error para buscar las características y umbrales adecuados

se puede evitar. El algoritmo SEaTH también ayudó a reducir al mínimo la participación humana en los pasos de clasificación y acelerar el proceso de clasificación cuando se utilizan imágenes de gran tamaño.

Como último punto, se investigó la contribución del IVM del sensor MODIS para mejorar el desempeño del análisis de imágenes basado en objetos. Este resultado pone de manifiesto que el suministro la información proporcionada por el IVM MODIS no sólo es importante para monitorear la fenología de los tipos de cobertura, sino también para diferenciar los tipos de cobertura que son difíciles de distinguir utilizando imágenes multiespectrales de una sola fecha.

El analysis de imágenes por objetos es visto como el nuevo paradigma para el análisis digital de imágenes. Sin embargo, no es perfecto. Todavía hay muchos aspectos que necesitan mayor desarrollo, entre ellos, la parametrización de la segmentación de imágenes y la evaluación de la exactitud de los resultados del análisis de imágenes basado en objetos, siendo dos temas importantes para su posterior estudio.

ABSTRACT

This research set out to discuss the effect of image segmentation to the classification accuracy and demonstrates through several applications how segmentation and object-based classification improve on pixel-based image classification methods, using different object-based classification methods over images of different spatial resolutions. A new algorithm SEaTH, which semi-automatically calculates the optimal features and thresholds for class differentiation, was tested in the land cover mapping and the MODIS Enhanced Vegetation Index (EVI) data was applied to test its contribution to the object-based image classification.

The core concept of object-based image analysis is that the important information necessary to interpret an image is not represented in single pixels, but in meaningful image objects and their mutual relationships. Two main reasons drive the development of object-based image analysis: available high spatial resolution imagery and improved computing tools; pixel based image analysis is limited.

Image segmentation creates image objects and it is the initial step in object-based image analysis. Only until recently image segmentation is implemented with a high level of precision and fast performance, which resulted in the development of object-based image analysis. By analyzing image objects, object-based image analysis is much closer to human vision and the classification results show both higher accuracy values and allow more detailed legends.

Image segmentation divides images into spectrally and spatially homogeneous objects. Evaluation of image segmentations before performing image classification ensures that the best segmentation result is used. By performing and evaluating several image segmentations with different parameter settings on Landsat-7 ETM+ imagery, the optimal segmentation result can be decided. By classifying the image with different segmentation results and evaluating the accuracies, it is found that best segmentations, in terms of objective function

rating, also led to the classifications with the highest accuracies, and, the accuracy values presented similar distribution as the objective function values in the function of the segmentations.

The performance of pixel-based and object-based image analysis was then compared over satellite images with different spatial resolutions: 10m, 30m, 100m, and 250m. Results showed that with the same training and reference data, object based image analysis obtained higher accuracy than that by pixel based methods with higher spatial resolution images among the test ones (10m and 30m); while, with the decreasing of spatial resolution, object-based image classification did not show higher accuracy. This experiment seems to suggest that object based image analysis has many advantages over the pixel-based one. While in the aspect of accuracy rating, the advantage only holds true for images with higher spatial resolutions.

Image objects can be characterized by a variety of spectral, spatial, texture, and contextual features. A comprehensive feature analysis is essential to work with image objects, and the trial and error process of searching proper features could hinder the utilization of the strength of object based image analysis. 'SEperability and THreshold' (SEaTH) algorithm is able to evaluate statistically any number of given features for any number of object classes of interest. This experiment with SEaTH algorithm shows that object based image analysis with features and thresholds identified by SEaTH algorithm produced a good classification result. With SEaTH, the time-consuming trial and error practice for seeking significant features and proper thresholds can be avoided, and thus SEaTH algorithm helped to minimize human involvement in classification steps and speed up the process of classification when huge datasets are to be dealt with.

As the last point, the contribution of MODIS EVI data to the improvement of OBIA with MODIS imagery was investigated. This result shows that the MODIS EVI data supply important information not only to monitor the phenology of the land cover types, but to differentiate land cover types which were difficult to be differentiated using multispectral image from a single date.

Object based image analysis is seen as the new paradigm for image analysis. However, it is by no means perfect. There are still many aspects that need further development, among them, image segmentation parameterization, and accuracy assessment for object based image analysis results, are two main topics for further study.

论文摘要

这项研究的目的是讨论影像分割对影像分类精度的影响，并且通过多种实验证明建立在影像分割的基础上的基于对象的影像分类比基于像素的图像分类有更多的优越性和准确性; 这些实验是使用多种基于像素和基于对象的分类方法并且使用有不同空间分辨率的影像。此外，这项研究也测试了一种新的算法 SEaTH 以及它对基于对象分类的贡献;这个算法可以半自动地计算出基于对象分类的最优 FEATURE 和阈值。同时，强化的植物指数 EVI 也用于基于对象的分类中来观察它对于基于对象的图像分类中分类精度提高的贡献。

基于对象的图像分析以对象为分类单元。对于基于对象的图像分析来说,其核心概念是: 重要的影像分析信息是不代表在单一像素中的，而是代表在有意义的图像对象及其相互关系中。驱动基于对象的图像分析的发展有两个主要原因: 1)：高空间分辨率航空影像逐渐变得普及, 可利用增强的计算工具;2) 基于像素的图像分析能力是有限的.

影像对象通常由图像分割产生，而且它是基于对象的图像分类的第一步。直到最近, 图像分割才得以准确和快速地实施，而这推动了基于对象的图像分析的进一步开发。通过分析图像对象，更接近人类视觉观察, 并且基于对象的图像分析结果即有较高的精度值又允许产生更详细的类别.

图像分割将影像分割成光谱上和空间上均匀的物体。在进行图像分类前, 评价图像分割的精度能够确保最佳的图像分割结果的运用。通过评价运用不同的参数设置的对 LANDSAT-7 ETM+ 图像进行分割的结果，可以选出最优影像分割结果。对影像的评价是基于客观方程的结果。通过对这些分割的影像进行分类和对分类结果的精度评价，发现最好的影像分割结果，根据客观方程的评价结果，也导致具有最高精度的影像分类结果，而且影像分类结果的精度值的分布和影像分割的客观方程值的分布相似.

运用具有不同空间分辨率的航空影像, 其空间分辨率是 10 米， 30 米， 100 米 和 250 米, 基于像素的影像分类和基于对象的影像分类结果进行了对比。结果表明，运用相同的训练和参考数据，对于具有较高的空间分辨率的影像（10 米 和 30 米）, 基于对象的图像分析获得比基于像素的分类方法更高的精度，而随着空间分辨率的下降，和基于像素的影像分类相比, 基于对象的影像分类并没有显示在精度上的优越性。这个实验似乎显示基于对象的影像分类有很多优势超基于像素的影像分类的优势。而在分类精度等级方面，这种优势仅适用于具有高空间分辨率的影像.

图像对象具有多种光谱，空间，质感，和上下文特征。为了确保影像分析结果的精度, 对于影像多种特征的全面分析和正确选择最有效的影像特征进行影像分

类是至关重要的，而寻找适当的影像特征的过程中可能会阻碍基于对象的影像分类的正确利用。在这一点上，建立在统计学的基础上，SEaTH 算法能够用于评价任何数目的影像特征并且计算出其域值。这个 SEaTH 算法能够利用广泛的对象特征因而为一个成功的影像分类提供了基础。这个基于对象的分类使用了由 SEaTH 算法计算出的对象特征和域值,并产生了很好的分类结果。使用 SEaTH 算法，可以避免因反复寻找合适的对象特征和正确的域值而延误的时间，SEaTH 算法，也有助于尽量减少在影像分类中人为的参与，并有助于在需要对庞大的数据加以处理时提高分类的速度和效率.

论文的最后一点，实验并讨论了增强的植物指数对使用 MODIS 数据的基于对象的分类的贡献。结果表明，增强的植物指数不但对监测物候的土地覆盖类型提供了重要信息，而且有助于区分那些用单日期多光谱影像的光谱信息难加以区分的土地覆盖类型.

基于对象的图像分析，被视为图像分析的新的范例。不过，目前它还处在发展中的阶段，有许多方面仍需要进一步发展。其中，图像分割参数的选取和图像分析结果精度评估是基于对象的影像分析中需要作进一步研究的两个主要议题。

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF SYMBOLS AND ACRONYMS

μm:	Micrometer
Arc/Info:	GIS software
AOI:	Area of Interest
DEM:	Digital Elevation Model
DN:	Digital Number
eCognition:	Object based image processing software developed in Munich, Germany
ENVI:	Image processing software
EM:	Electromagnetic
EMR:	Electro-magnetic Radiation
EO:	Earth Observation
ETM+:	Enhanced Thematic Mapper of Landsat-7
EVI:	Enhanced Vegetation Index
Fig:	Figure
GCP:	Ground Control Point
GIS:	Geographic Information System
GLCM:	Grey Level Co-occurrence Matrix
GSD:	Ground sampling distance
HS:	Human settlement
IFOV:	Instantaneous field of view
Imagine:	Image Processing Software
INEGI:	Instituto Nacional de Estadística Geografía e Informática
Ilwis:	Image Processing Software
J:	Jeffries-Matusita
Km^2:	Square Kilometers
Landsat-7:	American earth observing satellite
Lat:	Latitude
Lon:	Longitude
M:	Meter
MIR:	Middle Infrared part of the EMR
MLC:	Maximum Likelihood Classifier
MD:	Minimum Distance Classifier

MODIS:	Moderate spatial resolution Imaging Spectrometer
MS:	Multispectral
NDVI:	Normalized difference vegetation index
NIR:	Near Infrared part of the EMR
nm:	Nanometer
NASA:	National Aeronautics and Space Administration
NN:	Nearest Neighbour (classifier)
OA:	Overall accuracy
OBIA:	Object-based image analysis
PA:	Producer's Accuracy
PAN:	Panchromatic
PCA:	Principal Component Analysis
RMSE:	Root Mean Square Error
SEaTH:	SEparability and THreshold
SNR:	Signal to noise ration
SPOT:	Satellite Pour I'Observation de la Terre (France)
SPRING:	Object based image processing software developed in INPE, Brasil
SR:	Spatial Resolution
Tab:	Table
UA:	User's Accuracy
VI:	Vegetation Index
VNIR:	Visible and near Infrared

CHAPTER 1

INTRODUCTION

1.1. Problem formulation

Image analysis by visual interpretation of aerial photos or satellite imagery, can produce detailed land-cover thematic maps with higher accuracy than digital image classification (Mas *et al*. 1996, Sader 1992). However, this method is slow, expensive (in terms of time expenditure and expertise requirements) and subjective (since the performance of visual interpretation depends on the experience and knowledge of individual interpreters). Pixel-based digital image classification uses spectral information of single pixels which can lead to very sophisticated and detailed classification results if all classes of interest can be described unambiguously using spectral information only. However, for most applications this precondition is not given. Classification using pixel-based approaches very often shows the speckled appearance, which not only reduces the classification accuracy, but also complicates the further processing of classification results in a geographic information system (GIS) if applicable, since an even higher number of polygons (sometimes at the size of single pixels) is resulting from a very heterogeneous classification result (Ebert 2006).

Pixel-based digital image classification approaches use mainly spectral information of single pixels and sometimes texture information in certain environments (3*3 or 5*5 pixels, etc.), but not information related to shape (such as area, length, perimeter, etc.) of objects which corresponds to the

objects of interest, texture within the objects, neighbourhood, context, and levels. In most cases, information important for the understanding of an image is not represented in single pixels but in meaningful image objects and their mutual relations (Blaschke 2003). Human interpreters derive less information from the brightness of individual pixels than they do from the context and the patterns of brightness, of groups of pixels, and from the sizes, shapes, and arrangements of parcels of adjacent pixels (Campbell 2002, Lang *et al*. 2006). One way to make use of this additional information in digital image classification is to organize the image into objects that represent regions of similar pixels prior to the classification.

Image segmentation delineates units according to a certain criteria of homogeneity and, at the same time, requires spatial contingency (Lang *et al*. 2006). It is appealing for remote sensing applications because human vision tends to generalize images into homogeneous areas. Research into image segmentation is not new and several methods exist. They can be broadly categorized into measurement-space-guided spatial clustering, single-linkage region growing, spatial clustering, hybrid-linkage region growing, centroid-linkage region growing, and split-and-merge methods (Haralick and Shapiro 1985), or more simply, into edge-based and region-based algorithms (Blaschke and Strobl 2001). No matter which method is applied, segmentation provides building blocks for object-based image analysis. Image segmentation evaluation is a key issue, due to that segmentation produces the primitives of object-based image analysis and segmentation result has a direct effect on the following object-based image analysis. Image segmentation parameterization is still a topic under research. In this research, image segmentation, segmentation evaluation, and the effect of image segmentation to the image classification accuracy is one of the main topics.

Within an image object all kind of statistics based on single input layers or combinations within the input image layer stack can be computed such as spectral related information, texture within objects, shape, etc. Using image objects to calculate the statistics instead of pixels in a certain window improves the reliability of statistic without smearing edges, since objects do not exceed edges. Advantages of object-based analysis are meaningful statistic and texture

calculation, an increased uncorrelated feature space using shape and topological features, and the close relation between real-world objects and image objects. This relation improves the value of the final classification and can not be fulfilled by common, pixel-based approaches (Benz *et al*. 2004).

Object-based image analysis has the potential to use not only the spectral information, but the information regarding to the objects' shape, size, form, texture, and relations to the other objects from the super-and sub- levels, if a multi-spectral image segmentation and hierarchical classification is used. This extra information makes it a promising method for image analysis. This research carries out a comparative study between object-based and pixel-based image classification to fully explore the potential of object based classification approach in image analysis. Topics about the performance of OBIA such as with different spatial resolutions, using the nearest neighbour classifier and membership functions, using the semi-automatic feature extraction algorithm SEparability and Threshold (SEaTH, proposed by Nussbaum *et al*. 2006), and the contribution of Moderate spatial resolution Imaging Spectrometer (MODIS) Enhanced Vegetation Index (EVI) data will be covered in this research.

1.2 Hypothesis

The hypothesis of this thesis is that object based image analysis has advantage over pixel based one, and based on this, various researches were carried out to demonstrate how and on which grounds this advantage is based.

1.3 Research Questions

(1) What is image segmentation? How to evaluate image segmentation result?
(2) What is object-based image analysis?
(3) What is the impact of image segmentation to object-based image analysis? How to define the segmentation parameters to enable an object-based image analysis over-performs a conventional pixel-based one?
(4) How is the performance of object-based image analysis comparing to pixel-based image analysis? How object-based image analysis works with multi-spectral remote sensing images regarding to images with various spatial

resolutions?
(5) How object based image analysis result can be improved by semi-automatic feature extraction algorithm-SEparability and THreshold (SEaTH)?

1.4 Research Objectives

The objective of this research is to evaluate the contribution of image segmentation and object-based image analysis to land-cover classification using remote sensing images.

Based on a review of: image segmentation and segmentation result evaluation; object-based image analysis in land-cover classification in a forest area using remote sensing images; a comparison with conventional pixel-based image analysis if needs be, which supply the important antecedents for the thesis, the objective can be reached through the following sub-objectives:
(1) Research on segmentation optimization and its effect to object-based image analysis.
(2) Research on comparison of pixel-based and object-based image analysis over images with different spatial resolutions.
(3) Research on the object-based image analysis using a semi-automatic feature extraction algorithm SEaTH (proposed by Nussbaum *et al*. 2006).
(4) Research on the contribution of MODIS EVI data to the feature extraction with object based image analysis.

1.5 The outline of the thesis:

This thesis includes seven chapters, chapter 1, an introduction is given to this research, including the motivation, problem formulation, research questions and the objectives; chapter 2, theoretical background; chapter 3, Optimized image segmentation and its effect on classification accuracy; chapter 4, A comparison of the performance of pixel-based and object-based classifications over images with various spatial resolutions; chapter 5, A semi-automatic feature extraction algorithm and hierarchical network structure for object based image analysis; chapter 6, MODIS EVI data as an ancillary data for an object based land cover classification with MODIS data; chapter 7, conclusions.

CHAPTER 2

OBJECT-BASED IMAGE ANALYSIS

The purpose of this chapter is to give an introduction to the emerging field of object-based image analysis and to provide a comprehensive overview of methods involved and the respective background. Here I acknowledge the work of Lang *et al*. (2006), OBIA-Tutorial – Introduction to Object-based Image analysis, which provides the base for this chapter.

When performing digital image classification, our knowledge is often involved in the process and we want the computer imitate certain characteristics of the human way of image interpretation. In this aspect, in particular when going beyond spectral values of land cover classes, a pixel-based classification approach is limited, while an object-based approach supports considering spatial characteristics explicitly and addressing form-related as well as scale-related characteristics (Lang *et al*. 2006). However, the object-based approach also has its limitations and the human perception is still an ultimate benchmark, undefeated in analyzing complex scene contents with ease.

2.1. Image interpretation and perception

Human vision is well adapted for complex image interpretation tasks. We can hardly describe exactly what really happens if we look at an image and suddenly ‘see’ something. But indeed we notice that we do any kind of pattern recognition without major effort (Eysenck and Keane 1995, Lang *et al*. 2006).

However, human vision is challenged when dealing with remote sensing imagery (Lang 2005). There are three issues distinguishing interpretation of remote sensing imagery from interpretation conducted in everyday experience: 1) remote sensing images usually portray an unfamiliar perspective. 2) Many remote sensing images use radiation outside the visible portion of the spectrum. 3) Remote sensing images also portray the surface of the earth at unfamiliar scales and resolutions (Lang 2006). Experience is an important prerequisite for skillful and successful interpretation. Though through training, we can match imagery displayed in the false colour composite with natural phenomena and understand certain texture or structures and the imaged features, it does not prevent us from facing ambiguity when features are very like in structure or colour (Lang 2006). In this case, knowledge such as shape of and the spatial relations between the image regions may help in image interpretation (Blaschke 2003).

As shown in figure 2-1, we can distinguish the 'river' from 'lake' by its form, though they share similar spectral values. As for the 'municipal park', although it is spectrally similar to 'grassland', we can identify the parks using spatial form and context that are placed inside urban areas.

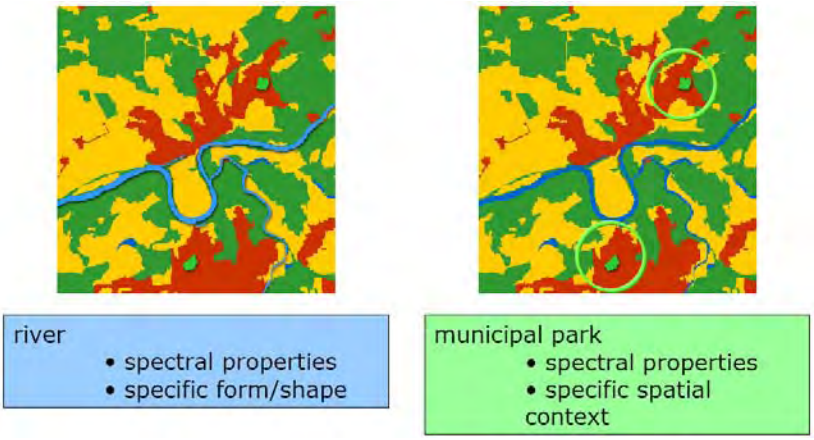


Figure 2-1: *Example of the way human interprets images using not only spectral features but form, shape, and context features. Source: Lang et al. 2006.*

2.1.1. Image data structure

A digital image is usually composed of many pixels which represent the brightness of regions on the earth's surface recorded digitally as numeric values (Campbell 2002). No matter which resolution, a pixel is always an averaged signal reflected or emitted by the observed underlying features.

Digital data stored in the remote sensing image can be viewed as a matrix. In the simplest case, the scene X can be described in the form of a matrix $X = (x)_{ij}$, where i = 1,…n; i is the image row dimension and j = 1,…m; i is the image column dimension. The pixel value in the image data is often referred as digital number (DN) and it is derived according to the spectral value reflected or emitted from the corresponding earth surface, and it is normally in the range of DN = 0, 1,…255 (in the case of 8 bit radiometric resolution, which is explained in section 2.1.2.3). Value 0 indicates the 'black' pixels meaning no reflected/emitted value, and the maximum absorption; value '255' indicates the white area in the image meaning maximum reflection and no absorption.

Multispectral image data with N spectral bands in the form of a matrix can be explained as $X = (x)_{ijn}$, with n=1, …, N spectral bands. One pixel in the multispectral N-dimensional image data can be described as

$$x_{ij} = (DN_1, DN_2, ..., DN_n)$$

In which DN_n is the digital value in the number-n spectral band.

2.1.2. The four dimensions of imagery

Imagery can be expressed in four dimensions: spatial, spectral, radiometric, and temporal. The following is the introduction of these aspects and the importance of each of the dimensions on information extraction from a scene.

2.1.2.1 Spatial resolution.

Spatial resolution is often expressed in terms of ground sampling distance (GSD) and refers to the area covered on the ground by a single pixel. Spatial resolution is based on various factors, such as the sensor field of view (FOV), altitude at which the sensor is flown, and the number of detectors in the sensor, etc. Furthermore, the spatial resolution of the sensors varies with the off-nadir angle and is also influenced by the terrain on the ground. Spatial resolution of satellite images is mainly influenced by the terrain differences and off-nadir viewing angles, as satellites usually have fixed orbit and a fixed spatial resolution at nadir. Table 2-1 summarizes the orbital height and associated spatial resolutions for some satellite images.

Table 2-1.
Spatial resolution of satellite sensor comparison

Satellite	Orbit height (km)	Spatial resolution
Landsat	700	30m MS, 15m PAN
SPOT 2, 4	832	20m MS, 10m PAN
SPOT 5	832	10m (VNIR), 20m SWIR, 2.5m PAN
IKONOS	681	4m MS, 1m PAN
Quickbird	450	2.44m MS, 0.61m PAN

As the spatial resolution increases, the associated file size increases. Table 2.2 shows the uncompressed file sizes of 4-band 8-bit imagery at various resolutions for a 1 km^2 area. Also the spatial resolution is related to swath width; the higher the resolution, the smaller the swath width of the image.

Table 2-2.
Spatial resolution versus file size

Spatial resolution (m)	30	15	10	5	2.5	1
File size (kb)	4	17	39	156	625	3906

2.1.2.2 Spectral resolution.

It refers to the number of spectral bands on a given sensor. Most of the aerial and satellite sensors take image in the visible and infrared regions of electromagnetic (EM) spectrum.

2.1.2.3 Radiometric resolution.

It is defined as the number of gray levels that can be recorded for a given pixel. A radiometric resolution of 8 bit will result in pixel values ranging from 0-255 and an 11 bit radiometric resolution can capture DN values ranging from 0-2047. The value range can be computed using the equation $N = 2^{R,}$, where N is the range and R is the radiometric depth.

2.1.2.4 Temporal resolution.

It refers to the time frequency when the same area of interest (AOI) is covered by the sensors. Image satellites are typically launched in a sun-synchronous orbit that results in the satellite revisiting a given spot on the earth at the same solar time. Further, sensors, such as IKONOS and Quickbird have the flexibility to take images off-nadir, increasing the frequency of the revisit. Also, at the higher latitudes, the frequency of revisits increases as compared to the equator.

2.1.3. Multi-spectral image analysis

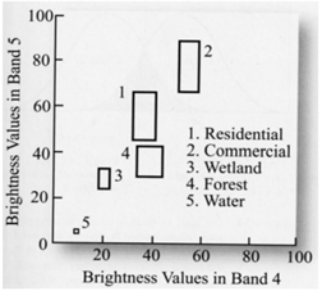
The overall objective of image classification procedure is to automatically categorize all pixels in an image into land-cover classes or themes. Normally, multispectral data are used to perform the classification and the spectral pattern present within the data for each pixel is used as the numerical basis for categorization. That is to say, different feature type manifests different combinations of digital numbers (DNs) based on their inherent spectral reflectance and emittance properties. A spectral “pattern” is a kind of spectral signature (SS), the better SS the higher spectral resolution and in this sense it is not at all geometric in character. Multispectral classification has found a great

deal of applications in agricultural assessment, forest management, urban planning, ecological monitoring, and water resources information collection. Several processing algorithms have been developed to perform this classification. In general, classification algorithms can be subdivided in two classes: (1) the parametric or statistical algorithms which assume a particular statistical distribution for each element to be classified and estimate the parameters of this distribution, such as mean vectors and covariance matrix, to be used in classification algorithms, and (2) the non-parametric algorithms, which are used when objects in the scene have reasonably distinct signatures. Nearest neighbour (NN) classifier is non-parametric which does not require the normal distribution of the sample pixels. While for other supervised classifications such as maximum likelihood (MLC) and minimum distance (MD) the distribution is assumed to be normal. The NN rule assigns an unclassified pixel to the same class as the nearest of n correctly classified samples. That is to say, given a collection of n reference points (in our case reflectance values of the pixel in each band), each classified by some external source, a new point is assigned to the same class as its nearest neighbour.

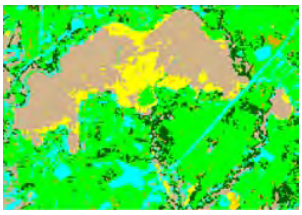
2.1.4. Pixel-based method

Typical pixel-based image classification classifies pixels. The classification feature is the spectral signature of the pixels represented by DN values (Figure 2-2).

Raw image



Feature space (Source: Jensen 2005).



Classified image

Figure 2-2: *Pixel-based classification process.*

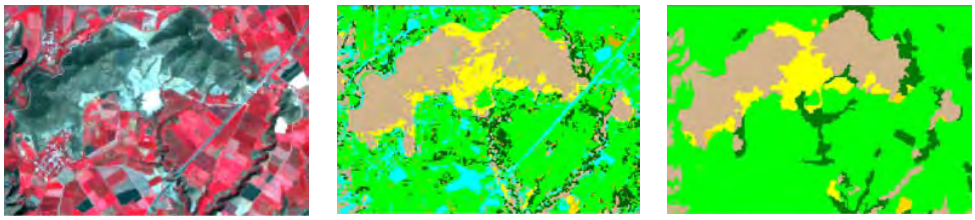
By comparing pixels to those of known identity (training samples), it is possible to assemble groups of similar pixels into information classes. These classes form regions on a map or an image, and they are, in theory, homogeneous: pixels within classes are spectrally more similar to one another than they are to pixels in other classes. In practice, each class displays diversity, as each scene exhibits some variability within classes (Campbell 2002, Lang *et al*. 2006). Spectral values belonging to more than one information class causes problem. The traditional pixel-based method assumes that pixels in the same land cover class have similar digital values and being close in spectral feature space. This does not hold true for complex environments and their respective classifications (Burnett and Blaschke 2003). Pixel-based analysis considers spectral reflectance values and texture in certain environment, but not information related to shape, neighbourhood, context, and levels. In most cases, information important for the understanding of an image is not represented in single pixels but in meaningful image objects and their mutual relations (Blaschke 2003). Human interpreters derive less information from the brightness of individual pixels than they do from the context and the patterns of brightness, of groups of pixels, and from the sizes, shapes, and arrangements of parcels of adjacent pixels (Campbell 2002, Lang *et al*. 2006). One way to make use of this additional information is to organize the image into objects that represent regions of similar pixels prior to the classification.

2.1.5. Object-based image analysis

Object-based image analysis classifies image objects. The basic elements of an object-based approach are image objects which are contiguous regions in an image and they are usually produced by image segmentation. The generation of image objects is suitable for textured or low-contrast image data. Image object primitives and objects of interest are different. Object primitives are usually the necessary intermediate step before objects of interest can be found by segmentation and classification process and only objects of interest match real-world objects. The smallest image object is one pixel. Image objects can be linked to a hierarchical network, where they are attributed with a high-dimensional feature space. Within an image object all kind of statistics based on single input layers or combinations within the input image layer stack can

be computed such as spectral related information, texture within objects, shape, etc. Using image objects to calculate the statistics instead of boxes of pixels improves the reliability of statistic without smearing edges, since objects do not exceed edges. Advantages of object-based analysis are meaningful statistic and texture calculation, an increased uncorrelated feature space using shape and topological features, and the close relation between real-world objects and image objects. This relation improves the value of the final classification and can not be fulfilled by common, pixel-based approaches (Benz *et al*. 2004).

Figure 2-3 illustrates the object-based classification results comparing with those by manual delineation and pixel-based classification. For many applications, e.g. land cover mapping, generalization is intrinsically required to produce tangible target objects of the same class which are relatively homogeneous according to the class definition (Blaschke 2003). The *salt and pepper* effect occurs if there are many pixels classified differently but actually belonging to the same land cover type, which produces an unnecessarily detailed classification of the land surface, and it can be overcome by object-based classification (Lang *et al*. 2006).



Original Landsat image | Pixel-based classification | Object-based classification

Figure 2-3: *A comparison of a Landsat image by pixel-based and object-based classifications.*

To extract objects of interest, the statistical analysis of pixels exclusively based on their spectral statistic may not be sufficient. The advent of higher resolution image data increased the need for more efficient methods more than ever. Generally, for high resolution data, segmentation as a pre-classification step is preferred over pixel-based classification because the resulting division of space tends to involve fewer and more compact sub-regions (Blaschke 2003).

Segmentation approaches are generally more suitable for high resolution data, where pixels tend to be spatially clumped (Blaschke 2003). For example, in a 1.0 m resolution image of a forest canopy, where each tree crown exhibits a 10 m diameter, each crown will be composed of many pixels. In this situation each 1.0 m pixel will be part of an individual crown. As a result, an image object tends to be composed of spatially clustered pixels that exhibit high spectral autocorrelation because they are all part of the same object. Consequently they have similar gray values. These characteristics correspond to Tobler's first law of Geography where 'objects are related to all other objects, but proximal objects are more likely to be related to each other' (Lang *et al*. 2006). In an image object, this relationship is both spectral and spatial (Hay *et al*. 2003).

2.2. Image segmentation

This part discusses segmentation algorithms to extract objects from imagery. The history of image segmentation is concisely introduced focusing on applications in remote sensing. The concept of image segmentation is also delivered highlighting the importance of homogeneity as a key criterion of the extracted objects. The practical aim of image segmentation is to find an optimum match between image objects and real world objects.

Segmentation algorithms may be categorized broadly into histogram-based, edge-based, and region-based. Histogram-based approaches perform segmentation within the feature space ignoring the spatial dimension in real world. It is a kind of unsupervised classification leading to classes but not to spatial space. Region-based algorithms deliver regions. They can be differentiated into region growing, region merging and splitting techniques and various derivations or combinations. Region growing starts with a set of seed pixels from which regions grow by adding neighbouring pixels as long as a "homogeneity criterion" applies. Here, the "homogeneity criterion" varies in different programs to perform image segmentation, and it can be calculated for example by the Euclidian distance. Region-merging starts with initial regions (single pixels) and merges them together until a scale-dependent threshold in size is reached. The splitting algorithms divide an image into regular sub-regions which again will be divided until a certain level of homogeneity is

reached. The combination of split and merge is realized in the split-and-merge algorithm, by which the image is subdivided into squares of a fixed size. Heterogeneous squares are subdivided again and homogeneous squares can be merged together. Edge-based algorithms search for edges that occur between homogeneous areas which usually includes filtering and enhancement of the image prior to the detection of the edges. The detected edges (groups of pixels) need to be combined in order to form a boundary (Lang *et al*. 2006).

2.2.1 The development of image segmentation

Although image segmentation techniques are well known in some areas of machine vision (Fu and Mui 1981, Haralick and Shapiro 1985) they are rarely used for the classification of earth observation data. One of the main reasons is that most of these algorithms were developed for the analysis of patterns, the delineation of discontinuities on materials or artificial surfaces, and quality control products. While the discrimination of earth observation imagery aims at the generation of spectrally homogeneous segments, which show the inherent dimensions/objects of the images (Blaschke *et al*. 2004). Common approaches use region growing or thresholding algorithms, but many derivatives for specific applications such as grey scale, hyper spectral images or data fusion of different sensors exist (Burnett and Blaschke 2003). As stated above, the idea of segmentation is not new but it is becoming widespread within the earth observation community recently. While the foundations of the basic principles were laid out in the 80ies (Haralick and Shapiro 1985) and various applications demonstrated the potential in the following years for environmental applications (Lobo *et al*. 1997), mainly the availability in commercial software packages catalyzed a boost of applications more recently (Blaschke *et al*. 2004).

2.2.2. Definition

Image segmentation divides images into homogeneous regions regarding to a certain criteria of homogeneity, and at the same time, requires spatial contingency (there are no space between two neighbouring objects) (Lang *et al.* 2006). Image segmentation is a preliminary step in object-based image analysis. By image segmentation, image is divided into homogeneous, continuous, and

contiguous objects. Image segmentation has been studied for a long time in the area of image analysis (Rosenfeld and Kak, 1976; Manjunath and Chellappa, 1991; Mao and Jain, 1992; Panjwani and Healey, 1995). But only recently, image segmentation and object-based image analysis are used in remote sensing image classification.

Pal and Pal (1993) defined image segmentation as a process of partitioning the image into some non-intersecting regions such that each region is homogeneous and the union of no two adjacent regions is homogeneous (Pal and Pal 1993). Formally it can be defined as follows: if F is the set of all pixels and P() is a uniformity (homogeneity) predicate defined on groups of connected pixels, then segmentation is a portioning of the set F into a set of connected subsets or regions (S_1, S_2, …, S_n) such that

$$\bigcup_{i=1}^{n} S_i = F \text{ with } S_i \cap S_j = \Phi, i \neq j \qquad (2.1)$$

The uniformity predicate $P(S_i) =$ true for all regions (S_i) and $P(S_i \cup S_j) =$ false, when S_i is adjacent to S_j. Note that this definition is applicable to all types of images.

There are hundreds or even more of the segmentation algorithms present in the literature, but there is no single method which can be considered good for all images, nor are all methods equally good for a particular image. Moreover, algorithms developed for one class of image may not always be applied to other classes of images. However, most of the segmentation methods developed for one class of images can be easily applied/extended to another class of images. Segments can be extracted in a variety of ways.

Image analysis implies to deal with image semantics (Lang *et al*. 2006). In most cases, important semantic information to understand an image is not represented in single pixels but in meaningful image objects and their mutual relations. Furthermore many types of image data are more or less textured. Airborne data, radar or very high resolution satellite data are playing an

increasing role in remote sensing. Analyzing such textured data can only be successful when they are segmented into meaningful homogeneous areas (Baatz and Schape 2000). Segmentation is the division of an image into spatially continuous, disjoint, and homogeneous regions. It is important that the internal heterogeneity of a segment under consideration is lower than that compared with its neighbouring areas (Blaschke *et al*. 2004).

2.2.3. Image segmentation methods

Image segmentation methods are commonly divided into three approaches: pixel-, edge-, and region-based (Pal and Pal 1993). Pixel-based methods include image thresholding and segmentation in the feature space. Edge based segmentation methods aim to find edges between regions and determine the segments as regions between the edges. Region based segmentation algorithms include region growing, splitting and merging techniques and their combinations (Blaschke *et al*. 2004). In this thesis, the segmentation methods used is region based.

1) Thresholding.

Some of the simplest approaches are all types of global thresholding. The spectral feature space is separated into subdivisions, and pixels of the same subdivision are merged when locally adjacent in the image data. Typically, this method leads to results of relatively limited quality. Over-segmentation and under-segmentation – i.e., separating into units which are too small or merging regions that do not belong to each other- take place easily without good control of meaningful thresholds. Local contrasts are not considered or not represented in a consistent way and the resulting regions can widely differ in size (Definiens 2004). Common alternatives are knowledge-based approaches, which incorporate knowledge derived from training areas or other sources into the segmentation process (Gorte 1998). These approaches mostly perform a pixel-based classification, based on clustering in a global feature space. Segments are produced implicitly after classification, simply by merging all adjacent pixels of the same class. In doing so, these approaches are typically not able to separate different units or objects of interest of the same

classification. Furthermore, the information on which classification can act typically is limited to spectral and filter-derivates (Definiens 2004).

2) Region growing

Region merging - ECHO (extraction and classification of homogeneous objects, after Kettig and Landgrebe, 1975) searches for spectrally similar neighbouring pixels and enlarges these groups to include adjacent pixels that have spectral values that resemble those of the core group (Lang *et al*. 2006). For example, the algorithm can first search for neighbourhoods of four contiguous pixels. For each group, it then tests members for homogeneity and pixels that are not similar to their neighbours are rejected from the group. Each of the homogeneous patches is then compared to each of its neighbours. If similar patches border each other, they are merged to form a larger patch. Patches are allowed to grow until they meet the edges of contrasting patches; when all patches reach their maximum extent within the constraints defined by the operator, the growing process stops. ECHO is a good example of a classifier that operates on fields of pixels rather than on each pixel in isolation. However, it performs the classification upon the average brightness of each patch, so it does not attempt to use image texture (Campbell 2002). Figure 2-4 is an example of region growing segmentation.

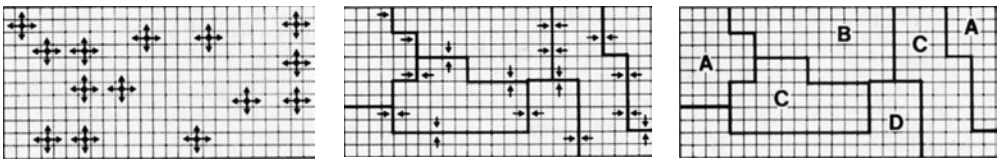


Figure 2-4: *Illustration of region growing image segmentation. Seed cells are distributed over image and 4-or 8-neighbourhood are included into region, if they do not belong to another region yet and if the homogeneity criterion applies; two neighbouring regions are unified if homogeneous criterion applies. Source: Campbell 2002, Lang et al. 2006.*

Region based segmentation starts from a set of given seed points and sometimes it suffers from lacking control over the breaking-off criterion for the growth of a region. Common applications of region based segmentation are

different types of texture segmentation algorithms (Gorte 1998, Blaschke *et al*. 2004, Lang *et al*. 2006).

3) Split and merge

Split and merge is one type of region-based segmentation. *Quadtree* is an example: initially image is seen as one object, then it is divided into 4 parts if homogeneity criterion does not apply, resulting in *quadtree* structure; then the homogeneous *quadtree* areas are merged.

Figure 2-5: *Quadtree segmentation. Source Lang et al. 2006.*

In region splitting and merging techniques the image is divided into sub-regions and these regions are merged or split based on their properties. In region merging the basic idea is to merge segments starting with initial regions. These initial regions may be single pixels or objects determined with help of any segmentation technique. In region splitting methods the input usually consists of large segments and these segments are divided into smaller units if the segments are not homogeneous enough. In an extreme case region splitting starts with the original image and proceeds by splitting it into n rectangular sub-images. The homogeneity of these rectangles is evaluated and each rectangle is subsequently divided into smaller regions until the homogeneity requirement is fulfilled. In both region merging and splitting techniques, the process is based on a high member of “pairwise” merges or splits. The segmentation process can be seen as a crystallization process with a big number of crystallization seeds. The requirement for the maintenance of a similar size/scale of all segments in a scene is to let segments grow in a simultaneous or simultaneous-like way (Blaschke *et al*. 2004, Lang *et al*. 2006).
Region based segmentation makes sense when images are composed of large,

compact and coherent objects and edge-based segmentation is better fitted for elongated structures (Lang *et al*. 2006). Here the edge signifies boundaries between homogeneous areas.

4) Edge based segmentation

Edge based segmentation is to partition an image based on abrupt changes-discontinuity- in gray level. An edge, in the image analysis literature, is a jump in intensity. The cross section of a so-called ideal edge has the shape of a ramp: infinite slope and flat portions on either side of the discontinuity (Wang 2005). To perform edge based segmentation, first, edges need to be detected, which generally involves process of filtering, enhancement, and detection. By filtering, image is smoothed and the noise in the image is decreased; by enhancement, local changes in intensities are revealed; and by detection, edge pixels are selected, for example choosing gaps by thresholding and combining and extending lines. After this first step, the detected edge pixels are linked to form the region boundaries.

Segmentation by representative measures is edge based and was proposed by Hoffman & Bohmer, 1999 (Lang *et al*. 2006). This method calculates a representative measure of each pixel for its neighbours, which forms the base of the algorithm. The values are calculated by a harmonic analysis of the spectral values for each spectral channel. The minima in the matrix of representative measure - typically arranged in pixel lineaments – represent spatial unsteadiness in the digital numbers. For the image segmentation, the vectorized minima of the representative measures delimit areas consisting of pixels with similar spectra properties (spatial segments) (Blaschke *et al*. 2004, Lang *et al*. 2006).

5) Watershed segmentation

A further relatively common procedure is watershed segmentation (Wegner *et al*. 1997). The name comes from the manner in which the algorithm segments regions into catchments basins. Typically, the procedure first transforms the original data into a gradient image, which can be considered as a topographic

surface. If this surface is flooded from its minima and if it is prevented the merging of the water coming from different sources, the image is partitioned into two different sets: the catchments basin and the watershed lines, and the catchments basin should theoretically correspond to the homogeneous grey level regions of this image. This method works for separating essentially convex and relatively smooth objects of interest that even may touch slightly in relatively homogeneous image data. When it works, it is convenient, fast and powerful. However, for remote sensing data, which typically contain a certain noise and not always strong contrasts, this method is typically not able to achieve appropriate results (Lang *et al*. 2006).

In Haris (1998), it is argued that watershed segmentation is a segmentation method with a high intuitive character and transparency. Spectral reflectance is modelled as height values and segments are built at gradient magnitudes along similar altitude levels, just in analogy of water flowing into valleys between watersheds. Region growing stops when neighbouring flooding regions meet each other. Higher scale segmentation is achieved by decreasing the number of local minima. One problem of watershed segmentation is that in an initial stage the algorithm leads to over-segmentation (Lang *et al*. 2006), and in many cases it has to be actively controlled by users. As the segmentation is only depending on spectral likeness, produced objects may vary significantly in size (Lang 2005).

6) Multi-resolution segmentation

Multi-resolution segmentation is the segmentation algorithm implemented in eCognition program (Definiens 2006). The design of multi-resolution segmentation is try to produce the segmentation results with strictly hierarchical multiple resolution, which convince the human perception and is reproducible. The segmentation speed should be satisfying (Lang *et al*. 2006).

Multi-resolution segmentation is based on region merging technique, and its homogeneity criterion includes both colour homogeneity and shape homogeneity. The shape homogeneity is represented by two more criteria of objects: compactness and smoothness.

The bottom up region merging technique starts with each pixel being a region. A pair of regions is merged into one region and each merging obeys a certain degree of fitting. Objects are merged into bigger objects as long as the merging is below a "least degree of fitting" (scale parameter) or as long as the merging fulfills the homogeneity criterion. The starting points for merging distributed with maximum distance and pairwise clustering process considers smallest growth of heterogeneity. The segmentation levels are established on several scales using different scale parameters (2^{th} level is based on the 1^{th} level, larger scale parameter results in larger image objects consisting of the objects of the lower level).

Image segmentation needs to address a certain scale parameter. However, most segmentation approaches do not allow the user to specify a certain scale of consideration and a level of detail or generalization, accordingly (Blaschke 2003). For example, the user does not usually specify information such as information of single bushes or trees or information of land cover units such as orchards. The flexibility in performing scale specific segmentation has led to a growing interest from landscape ecological applications of this multi-resolution segmentation (Lang *et al*. 2006). The hierarchical representation of process-relevant spatial units in various scale domains is one of the fundamental pillars within landscape ecology (Wu 1999, Lang *et al*. 2006). Image segmentation can be used to provide a consistent set of image primitives to be used as landscape objects (Burnett and Blaschke 2003, Lang and Langanke 2005).

About the *decision heuristics* of multi-resolution image segmentation, in the case of finding an adjacent object *B* to merge with an arbitrary object *A*, there are different degrees of fitting: 1) fitting: when the homogeneity criterion is fulfilled; 2) best fitting: when the homogeneity criterion is fulfilled, and the merge between *B* and *A* produces the best degree of fitting compared to the merge between A and any other adjacent object of *A*; 3) local mutually best fitting: find the best fitting object B for the object A, then find the best fitting object C for the object B. confirm that object C is the object A, otherwise take B for A and C for B and repeat the procedure (find the best fitting pair of objects in the local vicinity of A following the gradient of homogeneity); 4) global mutually best fitting: merge the pair of objects for which the

homogeneity criterion is fulfilled best in the whole image. 5) Distributed treatment order: use starting points with maximum distance to all other points treated before (treatment order defined over pixels or segments). 6) find the best fitting object B for the object A and confirm that the best fitting object C for the object B is indeed the object A (homogeneity criterion is fulfilled mutually). If not, take B for A and C for B and try again. The more detailed explanation about *decision heuristics* of multi-resolution image segmentation sees Lang *et al*. 2006.

In the definition of the degree of fitting, both colour and shape homogeneity are weighted against each other, and compactness and smoothness make up the shape homogeneity and are weighted against each other.

Colour homogeneity depicts the similarity between two objects in a certain feature space. As for the shape homogeneity, the criterion compactness: $h_{compactness} = \frac{L}{\sqrt{n}}$ depicts the relation between boundary length L of the object and the square root of the number n of the pixels of the object (square root of n equals the side of a square with n pixels). In the case of the ideal compact form of objects, objects do not become lengthy. The smoothness criterion: $h_{smoothness} = \frac{L}{b}$ depicts the relation between boundary length L of the object and the perimeter of the bounding box of the object (bounding box: shortest possible boundary length), and in the case of ideal smoothness the boundaries of the edges do not become fringed (Lang *et al*. 2006).

2.2.4. Image segmentation evaluation

Image segmentation is the preliminary and critical step in object based image analysis, and its proper evaluation ensures that the best segmentation result is used in the classification.

Overall visual survey is a commonly used method to evaluate the segmentation result. This step focuses on some general criterions, like the delineation of

varying land cover types (meadow/forest, agriculture/meadow, etc.), the segmentation of linear objects, the occurrence of faulty segmentations and a description of the overall segmentation quality.

Furthermore, detailed comparison based on visual delineated and clearly definable reference areas can also be carried out. Usually, different areas (varying in location, form, area, texture, contrast, land cover type etc.) are selected and each of them is visually and geometrically compared with the related part in the segmented image. The geometrical comparison is a combination of morphological features (area A_i, perimeter P_i, and shape Index SI_i) of the region *i* and the number of segments or partial segments in the case of over-segmentation (Neubert *et al*. 2006).

$$SI_i = \frac{P_i}{4\sqrt{A_i}}$$

the Shape Index comes from landscape ecology and address the polygon form. For all features the variances to the reference values can be calculated. For example, all polygons with at least 50% area in the reference object are counted in the case of partial segments.

Additionally, the quality of the objects can be visually rated (0 poor, 1 medium, 2 good). A good segmentation quality is reached when the overall differences of all criteria between the segmentation results and the associated reference objects are as low as possible.

2.3. Object-based image analysis

In this part, we introduce object-based image analysis. Since objects are aggregated by n-pixels we can use any feature statistical derivatives like mean, standard deviation, ratio, etc. Besides, there are a variety of geometrical features can be used such as shape, texture, etc. In addition, when working with hierarchical representations, there are features which characterize the relationships among objects on different hierarchical levels. All these features

can be used for classification. Using shape information, we can for example differentiate between *grassland* and a *football ground*. Using size information, we can tell small from big football grounds. Using spatial context information we can distinguish a big football ground within a city from a big football ground in the countryside. As this process has a strong modeling component, working with samples alone may be limited. Therefore it is recommended to focus on formulating rule sets and perform rule-based classification.

2.3.1. Definitions of object-based image analysis

Object-based image analysis is a sub-discipline of GIS science devoted to partitioning remote sensing imagery into meaningful image-objects, and assessing their characteristics through spatial, spectral and temporal scales. At the most fundamental level, object based image analysis requires image segmentation, attribution, classification and the ability to query and link individual objects in space and time. In order to achieve this, it incorporates knowledge from a vast array of discipline involved in the generation and use of geographic information (Hay and Castilla 2006).

By object-based image analysis, a number of objects are assigned to a certain class according to the class's description which describes the typical properties or conditions the desired classes have. The objects then become assigned according to whether they have or have not met these properties/conditions (Definiens 2004, Lang *et al*. 2006). When considering image objects instead of pixels, a wealth of additional features can be used for characterization. Both statistically aggregated spectral features and geometrical and neighbourhood-related features can be used, by which the dimensionality of the resulting feature space is getting significantly higher and rises to virtually limitless extent (Lang *et al*. 2006). If still a network of image objects is created, vertical relationships between super-objects and sub-objects in the object hierarchy can be used to expand the feature space even further (Lang 2005, Lang *et al*. 2006).

2.3.2. Sample- vs. rule-based classification

Image objects can be classified either by classifiers based on samples or using prior external knowledge stored in rule bases. By sample based classification, class membership of objects is defined by comparing similarity to selected samples. Samples are representative for their classes and the classification uses features which clearly distinguish the sampled class from other classes. Rather than delineating training areas, sample objects are iteratively selected. To ensure the accuracy of the classified images, these sample objects should have the most representative and clearly distinguished features (Lang 2005). In comparison to pixel - based training, the object - based approach of the nearest neighbour classifier requires less training samples/objects because one sample object already covers many typical pixel samples and their variations (Definiens 2004). Nearest neighbour classifier is sample-based by which object is assigned to the class whose sample objects are closest to it in the feature space.

Rule - based classification defines a class by rule(s)/membership functions using one feature or several features. The rule definition can be either fuzzy or crisp using object features or class-related features. Here the object features could be spectral related features such as mean, standard deviation, ratio, etc; shape features such as length, area, etc; texture features such as values of Grey Level Co-occurrence Matrix (GLCM). The class-related features could be features related to the neighbour objects or objects from the upper or lower level. Compared to sample-based classification, the rule - based classification can incorporate expert knowledge in the classification and formulate complex class descriptions. The classification process is also more transparent and the rule base is transferable (Lang *et al*. 2006). Integrating knowledge is a way to overcome the spectral similarity of different geographical features. Rules can use any of the spectral, spatial or hierarchical object features. Because of the enormous range of potential features, a rule-based approach needs formalized representation of the target class system and the way to interpret them beforehand (Lang 2005).

2.3.3. Fuzzy classification

Fuzzy set classification logic takes into account the heterogeneous and imprecise nature of the real world (Jensen 2005, p. 389). Fuzzy classification is a very powerful soft classifier which is used in neural networks (Gopal and Woodcock 1996) and probabilistic approaches (Curlander and Kober 1992, Lang *et al*. 2006). Avoiding arbitrary sharp thresholds, fuzzy logic is able to approximate real world in its complexity much better than the simplifying Boolean systems do. Fuzzy logic can model imprecise human thinking and represent linguistic rules (Benz *et al*. 2004). With fuzzy logic, it is possible to express each object's membership in more than just one class or the probability of belonging to other classes, with different degrees of membership or probabilities (Definiens 2004, Foody 2000, and Jensen 2005).

Fuzzy logic is a multi-valued logic quantifying uncertain statements (Lang *et al*. 2006). The basic idea is to replace the two boolean logical statements "true" and "false" by the continuous range of (0, 1) where 0 means "false" and 1 means "true" and all values between 0 and 1 represent a transition between true and false (Benz *et al*. 2004). "If - then" sentence is a common fuzzy rule to assign the objects into land cover class. If feature *x* (of the object) is member of the fuzzy set (associated with the class *forest*), the image object is a member of land-cover *forest*. The fuzzy sets can be combined to create the advanced fuzzy rules using operator "AND"- minimum operation; "OR"- maximum operation; "NOT"- inversion of a fuzzy value. Fuzzy rule-base (combination of the fuzzy rules of all classes) delivers a fuzzy classification. Every object has a set of values assigned to it with the degrees of membership to each class/degrees of class assignment. Since these values are possibilities to belong to a class, they don't have to add up to 1 (unlike probabilities). To produce results like maps for standard land - cover and land - use applications, the fuzzy results have to be translated back to a crisp value. To this end, the maximum membership degree of the fuzzy classification is used as crisp class assignment. This process is a typical approach for "de-fuzzification" of fuzzy classification results. If the maximum membership degree of a class is below a threshold, no classification is performed to ensure minimum reliability. As this output removes the rich measures of uncertainty of the fuzzy classification, this step

should be only performed if necessary and as late as possible in the whole information extraction process (Lang *et al*. 2006).

2.3.4. The strength of object-based image analysis

By object-based image analysis, additional information objects such as shape, texture, and relationship to other objects/semantic information can be used in the classification. By generating objects, the signal/noise ratio is increased and the number of units to be classified is decreased. By image segmentation, image resolution becomes adaptable. Since the classification is based on objects, the classification can be as simple as "click and classify". The object-based classification is a data base query. Any step and setting during the entire classification process is documented, and can be assessed and adopted if needed. Although the result is not necessarily more accurate-with higher accuracy, it can be reproduced and the process is to a high degree comprehensible. The formalized approach of analysis (the class definitions and composition and the documentation of the workflow and settings in the semi-automated process) technically allows for a transfer of the classification to other scenes (Lang and Langanke 2004, Benz *et al*. 2004, Lang and Langanke 2006, Lang *et al*. 2006).

2.4. Accuracy assessment

Accuracy assessment is an important part in image analysis, in order to have at hand an evaluation about the level of potential confusion, the reliability of the class assignments and the overall quality of the results. Here we discuss different approaches to accuracy assessment, starting with non-site-specific map comparison, then more advanced site-specific approaches with the error matrix as a core element, including *overall accuracy*, *user's*- and *producer's* accuracy as well as kappa statistic. Here we also elaborate on some problems related to the use of error matrix and problems that explicitly apply to object-based accuracy assessment. Especially the spatial geometrical characteristics of the extracted objects need to be evaluated which is a challenging task and which is still an open field of research.

2.4.1. Definitions

Remote sensing image classification accuracy can be defined as degree to which the derived image classification agrees with reality or conforms to the "truth" (reference map, assumed to be accurate, standard for the comparison); the error is the discrepancy between the situation depicted on the classified image and reality (Campbell 2002, Foody 2002). Accuracy assessment is important since it evaluates the usefulness of classified images/thematic maps for land management or the validity of maps for scientific investigations.

It is important that the quality of thematic maps derived from remote sensing data be assessed and expressed in a meaningful way, which not only provides a guide to the quality of a map and its fitness for a particular purpose, but also helps to understand error and its likely implications, especially if allowed to propagate through analyses linking the map to other data sets (Foody 2002). Remote sensing-derived thematic maps should normally be subjected to a thorough accuracy assessment before being used in scientific investigations and policy decisions (Stehman and Czaplewski 1998, Campbell 2002, Jensen 2005, Lang *et al*. 2006).

To correctly perform a classification accuracy assessment, it is necessary to systematically compare pixel or polygons in a remote sensing-derived classification map and ground reference information (which may in fact contain error) (Jensen 2005).

2.4.2. Non-site/site specific assessment and error matrix

Non-site specific accuracy assessment was based on comparisons of the areal extent of the classes in the derived thematic map relative to their extent in reference data set. A major problem is that the apparent and quantified accuracy of the map would hide its real quality (Foody 2002). This method is inaccurate in itself due to the possibility of compensating errors that don't show up and only overall figures of the two images are compared (Campbell 2002, Lang *et al*. 2006).

The standard form for reporting site-specific error is the error matrix, sometimes referred to as the confusion matrix because it identifies not only overall errors for each category but also misclassifications by category. Compilation of an error matrix is required for any serious study of accuracy. The error matrix consists of an n x n array, where n represents the number of categories (Campbell 2002). The intersection of the rows and columns summarize the number of sample units (pixels, clusters of pixels, or polygons) assigned to a particular category relative to the actual category as verified in the field. The diagonal of the matrix summarizes those pixels or polygons that were assigned to the correct class. Every error in the remote sensing classification relative to the ground reference information is summarized in the off-diagonal cells of the matrix. Each error is both an omission from the correct category and a commission to the wrong category. The column and row totals around the margin of the matrix (referred to as marginal) are used to compute errors of inclusion (commission errors) and errors of exclusion (omission errors). The outer row and column totals are used to compute the producer's and user's accuracy. Some recommend that the error matrix contain proportions rather than individual counts (Stehman and Czaplewski 1998, Jensen 2005, Lang et al. 2006). Presently, the error matrix is at the core of accuracy assessment but there is much scope to extend the analysis beyond it (Congalton 1994, Congalton and Green 1999, Foody 2002).

Table 2-3:
An example of a simple error matrix

Classification data	Reference data			
	Class 1	Class 2	Class 3	total
Class 1	6	2	1	9
Class 2	3	7	2	12
Class 3	0	1	8	9
total	9	10	11	30

The overall accuracy of the classification map is determined by dividing the total correct pixels (sum of the major diagonal) by the total number of pixels in the error matrix. It summarizes the total agreement between the classified

image and the reference image and only incorporates the major diagonal and excludes the omission and commission errors. In table 2-3, the overall accuracy: (6+7+8) / 30 = 70%

Producer's accuracy (Omission error) indicates the probability of a reference pixel being correctly classified. It is calculated by dividing the total number of pixels in that class with the number of correctly classified pixels. In table 2-3, the producer's accuracy of class 1: 6 / 9 = 67%

User's accuracy (Commission error) is the probability that a pixel classified on the map actually represents that category on the reference image. It is measured by dividing the total number of pixels that were actually classified in a class with the total number of correct pixels in that class. In table 2-3, the user's accuracy of class2: 7 / 12 = 58%.

2.4.3. Limitations of the error matrix and specifics of object-based accuracy assessment

For accuracy assessment, there is no single universally acceptable accuracy measure but instead a variety of indices, each sensitive to different features (Stehman 1997a). The design of an accuracy assessment has several essential elements including the definition of an appropriate sample size and sampling design. The sample size, for example, must be selected with care and be sufficient to provide a representative and meaningful basis for accuracy assessment (Hay 1979). If a probability-based measure of classification accuracy is to be used, it is essential that the reference data were acquired according to an appropriate sampling design (Hay 1979, Stehman 1997a, Stehman 1999b).

A variety of errors, due to for example the spectral similarity of land cover classes, are encountered in an image classification and interests are typically focused on accuracy of derived thematic map/classified image (Lang *et al*. 2006). Normally, information of the misclassified classes can be obtained from the error matrix. However, unfortunately other sources of error contribute to the pattern of misclassification depicted in the confusion matrix. Non-thematic

errors can be large and particular concern is the error due to mis-registration of the classified image with the reference data (Canters 1997, Czaplewski 1992, Muller *et al*. 1998, Stehman 1997a, Todd *et al*. 1980). In fact, the reference data are just another classification which may also contain error (Congalton and Green 1999, Khorram 1999, Lang *et al*. 2006). Problems with the accuracy of reference data may be particularly severe if a remote sensing data set is used as the reference data.

The erroneous allocations made by a classification are typically not randomly distributed over the region (Congalton 1988). Unfortunately, however, the confusion matrix and the accuracy indices derived from it provide no information on the spatial distribution of error. In classical accuracy assessments all misallocations are often equally weighted, while some errors are more important or damaging than others (Forbes 1995, Naesset 1996, Stehman 1999a). In many cases, the errors observed in a classification are between relatively similar classes and sometimes they may be important while other errors may be highly significant (Felix and Binney 1989, Foody 2000a, Steele *et al*. 1998, Townsend 2000). A further source of error associated with the use of a hard classifier that allocates each pixel to a single class is the assumption that the image is composed of pure pixels (Foody 1996).

Using an object-based classification approach requires to adapt existing accuracy assessment methods and develop new algorithms that evaluate the accuracy of object-related features explicitly. In Lang and Langanke (2006) it was proposed to generate random points within objects and check the labels against a ground data layer to assess the accuracy of a thematic layer generated by object based image analysis. A set of objects can be selected prior to the training process to be used as reference information as an alternative (Lang et al. 2006). In the case of smaller test areas with a limited number of larger objects, every single object could be assessed in terms of its label (Lang and Langanke 2006). However, by far, geometrical accuracy is harder to evaluate. Regarding to geometrical accuracy, classified image objects can be visually checked against manual delineation (Koch *et al*. 2003), but a quantitative assessment requires GIS overlay techniques.

There are difficulties in performing object-based accuracy assessment, which satisfies the needs/requirements as being discussed by Congalton and Green (1999). Two main reasons explain this: a) 100% geometrical fit between thematic image which is produced by object-based image analysis and reference data is usually not given due to different segmentation algorithms and other ways of feature delineation; b) the thematic classes are not mutually exclusive when using rule bases based on fuzzy logic, such as classifiers of nearest neighbour or membership function. In this sense, the accuracy assessment for object based image analysis is a matter of evaluation of both geometrical agreement and semantic agreement (Lang 2005, Lang *et al*. 2006).

References

Allen, T.F.H. and Starr, T.B., 1982, Hierarchy: Perspectives in Ecological Complexity. University of Chicago Press, Chicago, London.

Baatz, M., and A. Schape, 2000, Multiresolution segmentation: an optimization approach for high quality multi-scale image analysis. In: J. Strobl and T. Blaschke (eds). Angewandte geographische informationsverarbeitung, XII., pp. 12 - 23.

Blaschke, T., and J. Strobl, 2001, What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. GIS – Zeitschrift fur Geoinformationssysteme 14(6), pp 12 - 17.

Blaschke, T., 2003, Object-based contextual image classification built on image segmentation – IEEE proceedings, Washington DC, CD-ROM.

Blaschke, T., C. Burnett, and A. Pekkarinen, 2004, New contextual approaches using image segmentation for object-based classification. Remote Sensing Image Analysis: including the spatial domain. Kluver Academic Publishers, Dordrecht, pp. 211 - 236.

Burnett, C., and T. Blaschke, 2003, A multi-scale segmentation/object relationship modeling methodology for landscape analysis. Ecological Modelling 168(3), pp. 233 - 249.

Campbell, J.B., 2002, Introduction to remote sensing, 3rd ed. – New York, London.

Congalton, R.G., Green, K., Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Lweis Publishers, 1999.

Definiens, 2006. Definiens professional User Guide 5. Definiens AG, Munich.

Eysenck, M. W. & Keane, M. T. (1995) Chapter 4: Theories of perception, movement, and action. In *Cognitive Psychology: A Student's Handbook,* Lawrence Erlbaum, Hillsdale, USA.

Foody, G.M., 1996, Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data, International Journal of Remote Sensing, 17, 1317-1340.

Forman, R.T.T. 1995. Land Mosaics: The Ecology of Landscapes and Regions. Cambridge University Press, Cambridge, UK.

Gorte, B., 1998, Probabilistic segmentation of remotely sensed images. ITC Publication Series No. 63.

Haralick, R. M., and Shapiro, L. G., 1985, Image segmentation techniques. Computer Vision Graphics and Image Processing, 29, pp. 100 - 132.

Haris, K., Efstratiadis, S., Maglaveras, N., Katsaggelos, A., 1998. Hybrid image segmentation using watershed and fast region merging. IEEE Trans. Image process. 7(12), 1684-1699.

Hay, G., T. Blaschke, D. Marceau, and A. Bouchard, 2003, A comparison of three image object methods for the multiscale analysis of landscape structure. International Journal of Photogrammetry and Remote Sensing, 57, pp. 327-345.

Hay, G.J., and Gastilla, G., 2006, Object-based image analysis: strength, weakness, opportunities, and threats (SWOT). 1st International Conference on Object-based image analysis (OBIA 2006), 4-5, July, 2006, Salzburg, Austria.

Jensen, J.R., 2005, Introductory Digital Image Processing: A Remote Sensing Perspective, 3rd edition, Prentice Hall – Upper Saddle River, New Jersey, pp. 393 - 401.

Lang, S., F Albrecht & T Blaschke, 2006, OBIA-Tutorial – Introduction to Object-based Image Analysis, v 1.0 – Salzburg.

Lang, S., and T. Langanke, 2006, Object-based mapping and object-relationship modelling for land use classes and habitats. PFG – Photogrammetrie, Fernerkundung, Geoinformatik, 1, pp. 1-18.

Lang, S., 2005, Image objects and landscape objects – Interpretation, Hierarchical Representation and Significance – unpublished PhD thesis, Salzburg.

Lang, S., C. Burnett, & T. Blaschke, 2004, Multi-scale object-based image analysis – a key to the hierarchical organization of landscapes. Ekologia, supplement 23, pp. 1- 9.

Lang, S., 2002, Zur Anwengung des Holarchiekonzeptes bei der Generierung regionalisierter Segmentierungsebenen in hochstauflosenden Bilddaten. In: T Blaschke (ed.) Fernerkundung und GIS: Neue Sensoren – Innovative Methoden. Heidelberg, pp. 24 -33.

Lobo A., 1997, Image segmentation and discriminant analysis for the identification of land cover units in ecology. IEEE Transactions on Geoscience and Remote Sensing, 35, pp. 1136 - 1145.

Neubert, M., Herold, H., Meinel, G., 2006, Evaluation of remote sensing image segmentation quality-further results and concepts. 1st International Conference on Object-based image analysis (OBIA 2006), 4-5, July, 2006, Salzburg, Austria.

Pal, N. R. and Pal, S. K., 1993, A review on image segmentation techniques. Pattern Recognition, 26, pp. 1277 - 1294.

Wang, W., 2005, An edge based segmentation algorithm for rock fracture tracing, Proceedings of the Computer Graphics, Imaging and Vision: New Trends (CGIV'05) 0-7695-2392-7/05, IEEE Computer Society.

Wegner, S., Oswald, H. & Fleck, E. 1997. Segmentierung mit der Wasserscheidentransformation. Spektrum der Wissenschaft 6, pp. 113-115.

Wu, J., 1999, Hierarchy and scaling: extrapolating information along a scaling ladder. Canadian Journal of Remote Sensing, 25, pp. 367 - 380.

Wu, J., and O.L. Loucks, 1995, From balance-of-nature to hierarchical patch dynamics: A paradigm shift in ecology. Quaterly Review of Biology, 70, pp. 439 - 466.

CHAPTER 3

OPTIMIZED IMAGE SEGMENTATION AND ITS EFFECT ON CLASSIFICATION ACCURACY[1]

Abstract. Image segmentation is a preliminary and critical step in object based image analysis (OBIA). Its proper evaluation ensures that the best segmentation result is used in image classification. In this paper, image segmentations with nine different parameter settings were carried out with a multi-spectral Landsat imagery and the results were evaluated with an objective function that aims at maximizing homogeneity within segments and separability between neighbouring segments. The segmented images were classified into eight land cover classes by Maximum Likelihood Classifier (MLC) and the classification results were evaluated with independent ground data comprising 600 randomly distributed points. The accuracy assessment

[1] Based on the study of this chapter, the article "Optimal region growing segmentation and its effect on classification accuracy" with the authors: Yan Gao, Jean F. Mas, Norman Kerle, and Antonio Navarrete is in review process for publication in International Journal of Remote Sensing.

The content of this chapter was also presented as poster titled "Optimised image segmentation and its effects to the classification accuracy" in 5th International Symposium Spatial Data Quality 2007-modelling qualities in space and time, 13-15, June, 2007, ITC, Enschede, the Netherlands. This poster won the best poster prize.

results for the classifications based on the nine segmented images presented similar distribution as that of the objective function values for the nine segmentations; the optimal segmentations, i.e. with the highest objective function values, also resulted in the highest classification accuracies. This result shows that segmentation result has a direct effect on the classification accuracy and the objective function, not only worked on single band image as proved by Espindola *et al*. (2006) but also on multi-spectral imagery as tested in this paper, is indeed an effective way to determine the optimal segmentation parameters for the following classification. Pixel-based MLC was also carried out to compare with the segment-based classifications. The test of McNemar (z^2 = 3.96) showed that accuracy for the optimal- segmentation-led classification is significantly different than that obtained by the pixel based MLC result, with p= 0.05. This result shows that the classification based on the optimal segmentation obtained accuracy that is significantly higher than that of the pixel based classification.

Key words: Image segmentation, Objective function, Object based classification

3.1. Introduction

Land-cover mapping by visual interpretation of aerial photos or satellite imagery can produce detailed land-cover thematic maps with better accuracy than digital image classification (Sader and Winne 1992, Mas and Ramirez 1996). However, the method is slow, expensive (in terms of time expenditure and expertise requirements) and to some extent subjective. Pixel-based digital image classification uses spectral information of single pixels which can lead to very sophisticated and detailed classification results if all classes can be described unambiguously using spectral properties only. However, for most applications this precondition is not given. Classification using pixel-based approach very often shows the salt and pepper effect, which not only reduces the classification accuracy, but also complicates the further processing of classification results in a geographic information system (GIS) if applicable, since an even higher number of polygons (sometimes at the size of single pixels) is resulting from a very heterogeneous classification result (Ebert 2006).

Pixel-based digital image classification considers spectral reflectance values and texture in certain environment, but not information related to shape, neighbourhood, context, and levels. In most cases, information important for the understanding of an image is not represented in single pixels but in meaningful image objects and their mutual relations (Blaschke 2003). Human interpreters derive less information from the brightness of individual pixels than they do from the context and the patterns of brightness, of groups of pixels, and from the sizes, shapes, and arrangements of parcels of adjacent pixels (Campbell 2002, Lang *et al*. 2006). One way to make use of this additional information in digital image classification is to organize the image into objects that represent regions of similar pixels prior to the classification.

Image segmentation is a crucial initial step in object-based image classification, the assumption being that results directly affecting the performance of the subsequent classification. One principal point of concern here is the selection of segmentation parameters, which has conventionally been based on trial-and-error approaches (Flanders *et al*. 2003, Giada *et al*. 2003, Gitas *et al.* 2004). In literature on segmentation or object based image classification very little attention is paid to optimal object definition. Object definition by segmentation comprises both the choice of spectral bands to be considered and the setting of a heterogeneity threshold. The optimal object definition is here defined as the level of segmentation that results in the highest classification accuracy. Espindola *et al*. (2006) recently proposed an objective function to decide which parameter settings generate the best segmentation results, based on intrasegment homogeneity and intersegment separability. The method is robust as it utilizes the inherent characteristics of images: variance and spatial autocorrelation, which have not been considered in image segmentation evaluation before (Pal and Pal 1993, Evans *et al*. 2002, Benz 2004). Evaluation of image segmentation results with this objective function before performing image classification ensures that the best segmentation result is used.

In this paper, image segmentation was performed with the software SPRING (Câmara *et al*. 1996) with various parameter settings, and segmentation results were evaluated with this objective function to determine the optimal parameters. The ultimate aim is not so much segmentation optimization *per se*, but rather to

assess its actual benefit on the resulting classification. A multi-spectral Landsat-7 ETM+ image covering an area of 3,634 km^2 and including multiple land-cover types, such as irrigated and rain-fed agricultures, grassland, tropical dry forest, temperate forest, lava flow and urban area, was used. Segmented images were classified in order to test the hypothesis that the best segmentations also lead to the classifications with the highest accuracy. Since Espindola *et al*. 2006 tested the objective function with only a single band while multi-spectral images are often used to carry out the segmentation and classification, this paper intends to find out how the objective function works on multi-spectral image, and to guide future users of SPRING and similar packages in achieving optimal segmentation results that demonstrably lead to improved classification accuracies.

3.2. Study Area and data

3.2.1 Study area

The study area is located in Michoacán state, central west of México, covering an area of approximately 58*60 km^2, within the longitude of 19° 02’ N and 19° 36’ N, and latitude of 102° 00’ W and 102° 32’ W (figure 3-1). Due to the mountainous landscape, the altitude variations and the climate conditions, the predominant vegetation on and around the Tancítaro Mountain are dry shrubs, temperate sub-humid forests, cold mountainous shrubs, and mountainous grasslands. Pine, oak and fir are dominant forest species (Velazquez 1997). In the surroundings, there are avocado and rain-fed agriculture fields. Except avocado, the main perennial crops are apple, peach and pear. The main types of rain-fed agriculture are maize, beans, red and green tomatoes, potatoes, chili pepper, and sugar cane. Besides traditional agriculture activities, there are extensive cattle grazing and forest management (Pulido and Bocco 2003). This area belongs to one of the most important hydrologic systems of México known as “Balsas” basin (Fuentes-Junco, 2000). There are 14 catchments of hydrological system distributed in a centripetal radial pattern from the summit to the lowlands (Fuentes-Junco 2000). The predominant soil types are Ochric and Humic Andosols together with the Leptosols and Regosols (INEGI soil map, 1983).

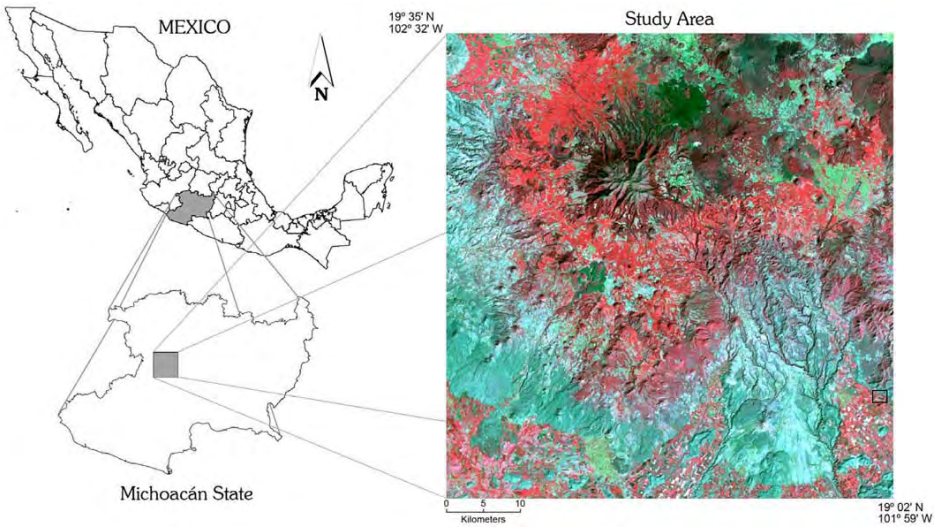


Figure 3-1*: The study area. Left side of the figure are two sketch maps indicating Mexico and Michoacán state where the study area is locating; right side is the false colour composite of Landsat image with red, green, and blue bands of 4, 5, and 7. The small box in the south-east corner indicates the area in figure 3-3.*

3.2.2. Land cover classes

There are eight land cover classes in the classification and they are defined in table 3-1. The rationale behind the development of the classes was: the classes are well adapted to the type of landscapes existent in regions with characteristics similar to the Mexican mainland; they are compatible with established ones (e.g, land use/cover map (2000) from forest inventory of México) in order to make it possible the comparison between our results and others using different land cover class nomenclatures; and the class definition matches the spatial resolution of used satellite imagery. In addition, the development of the land cover classes attended to the suggestion endorsed by Loveland *et al*. (2000), i.e. each vegetated land cover class represents relatively homogeneous land cover characteristics (e.g. similar floristic and physiognomic characteristics), that exhibit distinctive phenology (i.e. onset, peak and seasonal duration of greenness) and have comparable levels of relative primary production (Carrao *et al*. 2007).

Table 3-1:
Land-cover classes description and representative number of collected samples.

Class	Description	Number of samples
Human Settlement	Built-up areas and transport network.	59
Irrigated Agriculture	Agricultural areas irrigated artificially and periodically.	58
Rainfed Agriculture	Agricultural areas that are not artificially irrigated and consequently do not present vigorous vegetation during dry season.	41
Temperate Forest	Wooded areas where pine, oak, fir, cedar, etc. species predominate, which presents vigorous vegetation even during day season.	183
Tropical Dry Forest	Perennial and sub-perennial tropical forests including tropica evergreen and sub-evergreen forest from medium height to low height. It does not present vigorous vegetation during dry season.	131
Grassland	Natural or artificially areas with herbaceous vegetation.	32
Orchards	Artificially irrigated agricultural areas with mainly avocado trees.	42
Lava flow	Natural areas covered with lavas from the 1946 volcanic exploration of volcano Paricutin.	54

*Referenced on the classification scheme of the INEGI land use and vegetation cartography for natural land cover categories; there is no water bodies in the study area.

3.2.3 Earth Observation data

The available data comprise of a Landsat ETM+ image obtained on 16 February 2003 during the dry season, containing 6 bands with a spatial resolution of 30m; a mosaic of 25 ortho-corrected photographs taken in 1995 with 2 meters spatial resolution from Instituto Nacional de Estadística Geografía e Informática (INEGI), and a land-cover map generated from a

project “National Forest Inventory of México” in 2000.

Landsat-7 ETM+ data is the main source of remote sensing information used in this chapter (thesis). Reasons for this are its the broad spatial coverage, the sensors spatial resolution of up to 15 meters and its multiple spectral channels, including a panchromatic band. Although spectral bands are not overwhelming in number they are evenly distributed and well studied. The long term acquisition plan of the Landsat program predestines Landsat data for long term monitoring. The recording of land- and water surfaces started in 1972 with Landsat-1, therefore offering now 30 years of coverage (Arvidson *et al.* 2001, Goward *et al.* 2001, Flynn *et al.* 2001).

Landsat-7 was launched on April 15, 1999. The earth observing instrument onboard this spacecraft is ETM+ (Lillesand and Kiefer, 2000). ETM+ is a passive sensor that measures the solar radiation that reflected or emitted from earth surface. Landsat-7 collects data from nominal 705 kilometers altitude and 183 kilometers of swath with a repeat cycle of 16 days. The platform on Landsat-7 ETM+ is near-polar orbiting system sensor, its latitude cover is: 81 degrees north to 81 degrees south; its longitude cover is: 180 degrees east and west. Landsat-7 ETM+ data are recorded in 7 bands with spatial resolution of 15meters, 30 meters and 60 meters. The ETM+ instrument on the Landsat-7 spacecraft contains sensors to detect earth scene radiation in three specific bands: visible and near infrared (VNIR) bands - bands 1 (blue), 2 (green), 3 (red), 4 (near infrared), and 8 (Panchromatic) with a spectral range between 0.4 and 1.0 micrometer (um); short wavelength infrared (SWIR) bands - bands 5 and 7 with a spectral range between 1.0 and 3.0 um; and thermal long wavelength infrared (LWIR) band - band 6 with a spectral range between 8.0 and 12.0 μm.

Landsat-7 ETM+ imagery was geometrically corrected by 86 evenly distributed ground control points (GCPs) extracted from a mosaic of ortho-photographs with 2m spatial resolutions. The rectification was carried out in a sub-pixel level and the Root Mean Square (RMS) error was less than a pixel (16.8m). The image was re-sampled by Nearest Neighbour and was geo-referenced to UTM zone 13, with the spheroid and datum in WGS 84.

3.3. Methods

3.3.1. Image segmentation and region growing in SPRING

Image segmentation divides images into continuous and contiguous homogeneous regions. Such techniques can be grouped into three main types: thresholding/clustering, region based, and edge based. More information can be found in Fu and Mui (1981), Haralick and Shapiro (1985), and Pal and Pal (1993). Region growing techniques are being widely used for remote sensing applications and they guarantee creating closed regions (Espindola *et al.* 2006). In region growing, segments are formed starting from suitable initial pixels (seeds) by iteratively augmenting them with neighbouring pixels that satisfy a chosen homogeneity criteria. There are different ways to initiate the region growing process, for example using seed pixels (Stuckens 2000), seed areas (Adams and Bischof 1984), or initial edge detection-based segmentations (Egawa and Kusaka 1988). The process stops when all pixels are segmented into objects. Every image pixel is assigned to a segment and the image is completely filled with objects.

SPRING, a non-commercial programme, uses a region growing segmentation algorithm, which ranked second in segmentation quality among seven algorithms tested by Meinel and Neubert (2004). In their research, the commercial software eCognition obtained the best result. eCognition can derive a large number of particularly geometrical segment attributes, and thus shows a great advantage over traditional spectral-information based methods, mainly because of the integrated classification that takes advantage of those attributes. However, it is also expensive, in terms of licensing costs, hardware requirements and high user knowledge level, and is not affordable for many institutions. Thus substantial research has been focusing on the use of freely accessible software, such as SPRING.

The segmentation algorithm in SPRING uses region growing segmentation method which starts from suitable initial pixels (seeds) and augments them iteratively with neighbouring pixels that satisfy some homogeneity criteria. SPRING allows two algorithms for segmentation: "Region growing" and "basin detection". The first is a pixel merging technique, second a watershed

algorithm (edge detection, based on a Sobel-Filter, 3x3 Pixel Kernel). By testing both algorithms, "basin detection" leads to very poor results. “Region growing” algorithm in SPRING is based on the traditional region growing technique, with modifications which partially solve the problem of the dependence on the order of the merges (Bins *et al*. 1996). Image segmentation in SPRING uses two parameters: a similarity threshold and an area threshold to guide the segmentation procedure. “Similarity” is a threshold value that determines if two neighbouring pixels (objects) are grouped, while the “area” threshold is used to filter out the objects smaller than this value. It starts by comparing neighbouring pixels and merging them into regions if they are similar. The regions then try iteratively to merge the resulting regions. Two neighbouring regions, R_i and R_j are merged if they satisfy the following conditions:

(1) Threshold condition: $dist(R_i, R_j) \leq T$

(2) Neighbourhood condition 1: $R_j \in N(R_i)$ and $dist(R_{j,} R_i) \leq dist(R_k, R_i), R_k \in N(R_i)$

(3) Neighbourhood condition 2: $R_i \in N(R_j)$ and $dist(R_{j,} R_j) \leq dist(R_k, R_j), R_k \in N(R_j)$

In the above, T is the chosen similarity threshold, $dist(R_{i,} R_j)$ is the Euclidian distance between the mean grey levels of the regions and $N(R)$ is the set of neighbouring regions of region R. In addition, regions smaller than the chosen area threshold are removed by merging them with their most similar neighbour (Bins *et al*. 1996, Espindola *et al*. 2006).

Only very similar regions are merged first. The similarity threshold value must be manually provided by the user and, therefore, a tradeoff is inevitable: if it is set too low the growing process will generate over-segmented regions, otherwise segments representing different land-cover will be incorrectly merged together. The choice of this threshold value as well as the area threshold will greatly depend on the specific application and data (Câmara *et al.*

1996, Bins *et al.*1996, and Espindola *et al*. 2006). SPRING uses two parameters, "similarity" and "area", to guide the segmentation procedure. "Similarity" is a threshold value that determines if two neighbouring pixels (objects) are grouped, while the "area" threshold is used to filter out the objects smaller than this value. Segmentation results in SPRING depend on these two threshold parameters, and vary with the number of spectral bands utilized. Low values of similarity and area threshold result in excessive partitioning and over-segmentation, while high values force the union of spectrally distinct regions and result in under-segmentation. The same similarity threshold value results in larger segments for single spectral band images than for multiple band data, since the similarity threshold is based on the Euclidean Distance, which is calculated by the following equation:

$$D = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \qquad (3.1)$$

where D is the distance between two points p and q, and i is the dimension ranging from 1 to n. Euclidean Distance is larger in multiple dimensions than in one dimension. Thus, for a certain similarity threshold there are fewer pixels (objects) that can be merged, and thus the produced objects are smaller in multiple band image segmentation. However, multiple band images are also better able to differentiate objects, as different spectral bands complement one another in segmentation.

The mapping standard from National Cartographic Center of México (INEGE) can be a reference to define segmentation parameters. According to INEGE, the minimum mapping unit is 2*2 mm on the map, taking the most commonly used map scale of 1:100000, the minimum mapping unit in reality is 200m * 200m, by which in Landsat-7 ETM+ images (with 30 meters resolution), is approximately 6*6 = 36 pixels. So if the minimum area parameter in the segmentation is smaller than 36 pixels, this area won't be presented in a land-cover thematic map in a scale of 1:100000. While, from another point of view, if the parameter of "area" is given big value, then in the segmentation, two different land-cover types in small area can actually be grouped together in an

object, whose spectral information does not represent neither of the two land-cover types. In order to keep the pure spectral information from segmentation, the parameter "area" can first be set to a small value, and after the object-based image classification, the results can be processed with a kind of filter, for example, to make sure the minimum objects can have an area equal or more than 6*6pixels.

3.3.2. Evaluation of segmentation quality

The accuracy of segmentation directly affects the performance of object based image analysis. Only good segmentation result can lead OBIA over-perform pixel based classification. Pal and Pal (1993) claimed that a human being is the best judge to evaluate the segmentation output. In Benz 2004, human interpretation and correction was also listed as a method of validation of segmentation result. However, some attempts were made for quantitative evaluation. A two dimensional distance was used to measure the difference between a human being proposed segmentation and a segmentation done by an algorithm. Some performance parameters were also defined, such as region uniformity, region contrast, line contrast, etc. Details are in Pal and Pal (1993). In Evans *et al*. 2002, segmentation result was evaluated by quantifying the discrepancy between the segmentation result and the "ideal segmentation" which was produced manually. Methods were given to measure the degree of over-and under-segmentation of regions, and to measure the discrepancy between the positions of the region boundaries. Recently, Espindola *et al.* 2006 proposed an objective function to evaluate the segmentation results and experiment results showed that it is an effective method to decide the optimal segmentation for the following object based image analysis.

The function aims at maximizing homogeneity within segments and separability between neighbouring segments, and is thus based on two corresponding parameters employing variance and spatial autocorrelation. Intrasegment homogeneity is measured by calculating the variance of the regions produced by a segmentation algorithm, by the formula:

$$v = \frac{\sum_{i=1}^{n} a_i . v_i}{\sum_{i=1}^{n} a_i} \qquad (3.2)$$

where v_i is the variance of a segment and a_i is its area. The intrasegment variance v is an average weighted by the area of each region, which puts more weight on larger regions, avoiding possible instabilities caused by smaller ones.

Spatial autocorrelation is a well-known property of spatial data. Similar values for a variable tend to occur in nearby locations, leading to spatial clusters. By measuring spatial association, intersegment heterogeneity can be assessed. Moran's I autocorrelation index measures the degree of spatial association as reflected in the data set as a whole. The algorithm for computing Moran's I index (the spatial autocorrelation of a segment) uses the fact that region-growing algorithms generate closed regions. For each region, the algorithm calculates its mean grey value and determines the autocorrelation for all adjacent regions. In this case, Moran's I is expressed as:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - y)(y_j - y)}{(\sum_{i=1}^{n}(y_i - y)^2)(\sum_{i \neq j}\sum w_{ij})} \qquad (3.3)$$

where n is the total number of regions, w_{ij} is a measure of the spatial proximity, y_i is the mean grey value of region R_i, and y is the mean grey value of the image. Each weight w_{ij} is a measure of the spatial adjacency of regions R_i and R_j. If regions R_i and R_j are adjacent, w_{ij} = 1. Otherwise, w_{ij} = 0. Thus, Moran's I applied to segmented images will capture how, on average, the mean values of each region differs from the mean values of its neighbours. Small values of Moran's I indicate low spatial autocorrelation, in which case the neighbouring regions are statistically different. Local minima of this index correspond to locations of large intersegment heterogeneity. Such minima are

associated with segmentation results that show clear boundaries between regions. In the case of multi-spectral images, the Moran's I was calculated by averaging the values obtained for each band.

The proper choice of segmentation parameters combines a low intersegment Moran's I index (adjacent regions are dissimilar) with a low intrasegment variance (each region is homogeneous). The proposed objective function combines the variance measure and the autocorrelation measure in an objective function given by:

$$F(v, I) = F(v) + F(I) \qquad (3.4)$$

Function $F(v)$ and $F(I)$ are normalized functions, given by:

$$F(v) = \frac{v_{\max} - v}{v_{\max} - v_{\min}} \qquad (3.5)$$

and

$$F(I) = \frac{I_{\max} - I}{I_{\max} - I_{\min}} \qquad (3.6)$$

Here, $v_{\max}$ and $v_{\min}$ are the largest and smallest variance values in a group of experimented segmentations, respectively, and v is the variance value of the segmentation in evaluation. $I_{\max}$ and $I_{\min}$ are the largest and smallest Moran's I index values, respectively, in the group of experimented segmentations; I is the Moran's I index of the segmentation in evaluation. By these two functions, the normalized variance $F(v)$ and Moran's I index $F(I)$ are in the range of (0, 1), and consequently, the objective function $F(v, I)$ is in the range of (0, 2).

To calculate the intrasegment variance index, segments of each image were first exported as shapefiles from SPRING to Arc/info, where the variance of

each segment was calculated. For the spatial autocorrelation index, the mean grey values of each segment and the complete image were also obtained through calculations in Arc/info.

3.3.3. Segmentation-based classification, accuracy assessment, and the test of McNemar

After segmentation, the average spectral response of each segment was calculated and a supervised image classification was performed to those homogeneous regions. In Maximum Likelihood Classification (MLC) the assumption is that the training samples for each land-cover class are normally distributed. We put mean spectral information in the segments to form segmented images in which pixels in each segment have the same spectral information. A method is utilized that allows MLC to be used to classify both pixel-based and segmented images (figure 3-2). The basic approach is to generate training data based on the pixel-based image to take into account the variability of the spectral response of each land-cover class, based on the statistic value of the generated training samples, the original pixel-based image and the segmented images are classified by MLC.

For each of the eight land cover classes, representative ground reference data were collected and used for training and testing purpose of the study. Homogeneous sample units, each corresponding to a Landsat pixel area, were selected for each land cover class using land use map (2000) as strata. High spatial resolution earth observation data, namely ortho-corrected aerial photographs from 1995, covering the whole study area, were used as the base data source to recheck reference land cover classification of sample units. Reference information for each sample unit was derived by visual interpretation of high resolution earth observation data overlaid with a co-registered 30m data grid corresponding to the Landsat imagery. In the sample data collection process, we spread as much as possible each class samples over the entire study area, in order to take into account possible regional differences,

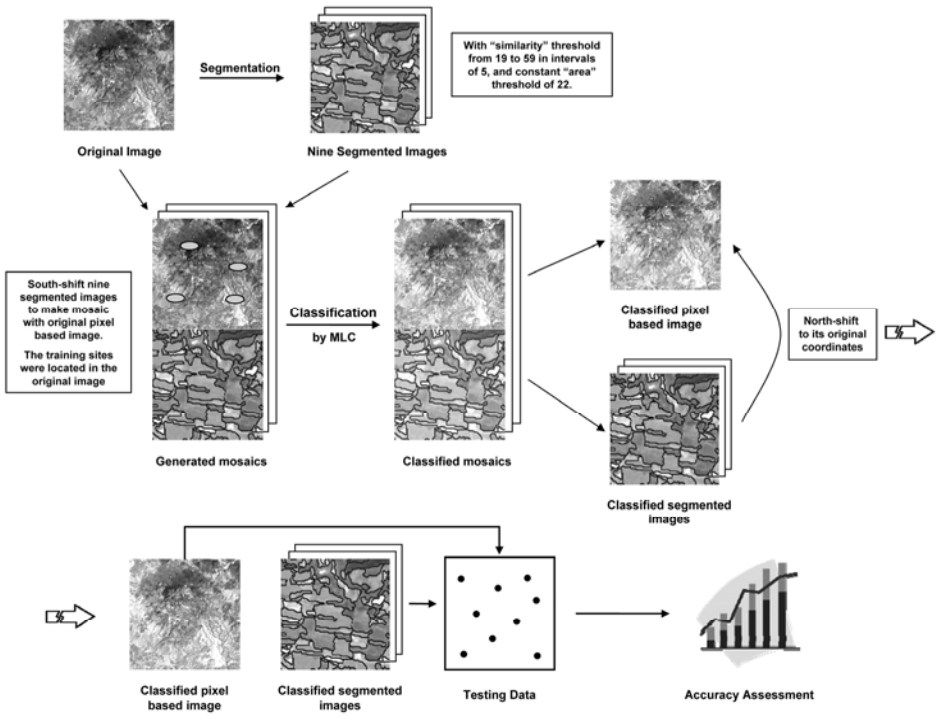


Figure 3-2: *illustration of proposed segment based image classification procedure.*

and to prevent geographic correlation due to adjacent pixels (Hammond and Verbyla 1996). Although it is open to debate the appropriate number of sample units per class that are needed to train a supervised classifier, the minimum sample units are related to the used algorithm itself; to the feature space dimension; to the sample selection method, and to the size and spatial variability of the study area (Carrao *et al*. 2007, Foody and Mathur 2006, Foody *et al*. 2006, Huang *et al*. 2002, Jensen 1996, Mather 2004). However, certain classification problems have non-zero Bayes error, no matter the sample size or the feature space dimension (Ho and Basu 2002).

Accuracy assessment was carried out using 600 stratified random points. 75 random points were put to each of the eight classes of interest in a land use/cover map (2000). These points were interpreted based on the ortho-

corrected photographs, land use map 2000, data from field survey, and Landsat-7 ETM+ image. Error matrices were generated and standard accuracy indices such as the Overall accuracy, User's and Producer's accuracy were calculated (Boschetti *et al*. 2004).

Map accuracy statements are often compared to evaluate the relative suitability of different classification techniques for mapping. Accuracy statements should be compared in a statistically rigorous fashion to provide a more objective basis for comment and interpretation (Foody 2004). In many remote sensing studies, the same set of ground data are used in the assessment of the accuracy of the thematic maps to be compared. For related samples, the statistical significance of the difference between two accuracy statements maybe evaluated using tests that take into account the lack of independence such as The test of McNemar. It is a non-parametric test that is based on confusion matrixes that are 2 by 2 in dimension (table 3-2).

Table 3-2:
The definition of Matrix Elements used in Equation 3.7.

	classification 1	
Allocation	correct	incorrect
classification 2		
correct	f_{11}	f_{12}
incorrect	f_{21}	f_{22}

The attention is focused on the binary distinction between correct and incorrect class allocations. The McNemar test equation can be expressed as

$$Z^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \qquad (3.7)$$

In which f_{ij} indicates the frequency of verification samples lying in confusion

matrix element i and j. f_{12} and f_{21} are the number of pixels that with one method were correctly classified, while with the other one were incorrectly classified. Z^2 follows a chi-squared distribution with one degree of freedom (Agresti 1996). The statistical significance can be obtained with the derived z^2 compared against tabulated chi-square values. For example, with one degree of freedom, when the calculated z^2 >= 3.84, the two classifications are significantly different at 0.05% level; when z^2 >= 6.64, the two classifications are significantly different at 0.01% level; when z^2 >= 10.83, the two classification are significantly different at 0.001% level; and when z^2 < 3.84, the two classifications are not significantly different.

3.4. Results and Discussions

3.4.1. Image segmentation results

The Landsat-7 ETM+ image was segmented, whereby experiments showed that for this imagery, in this study area, and for the expected land cover types, segmentation results varies more significantly with varying "similarity" threshold than with varying "area" threshold, and thus it appears that the "similarity" threshold has a more significant influence on the segmentation results than "area" threshold. Nine segmentations were generated with various parameter settings. The tested parameter settings use "similarity" thresholds ranging from 19 to 59[*] in intervals of 5. The selecting of "similarity" threshold was based on visual evaluation of the segmentation results. With value 19, it appeared that the image was over-segmented and with value 59 under-segmented (Figure 3-3). So the tested "similarity" value was set between 19 and 59. The "area" threshold was kept at a constant 22 in accordance with recommendations by Espindola *et al*. (2006) who used a Landsat image and found optimal segmentation results with this "area" threshold. The segmentation results are presented in figure 3-3, which shows that the bigger the "similarity" value, the larger the generated segments are. Note that for large images, the segmentation procedure can be time-consuming.

[*] Here the values are chosen based on trial and error practices, the higher this similarity value, the bigger the generated objects.

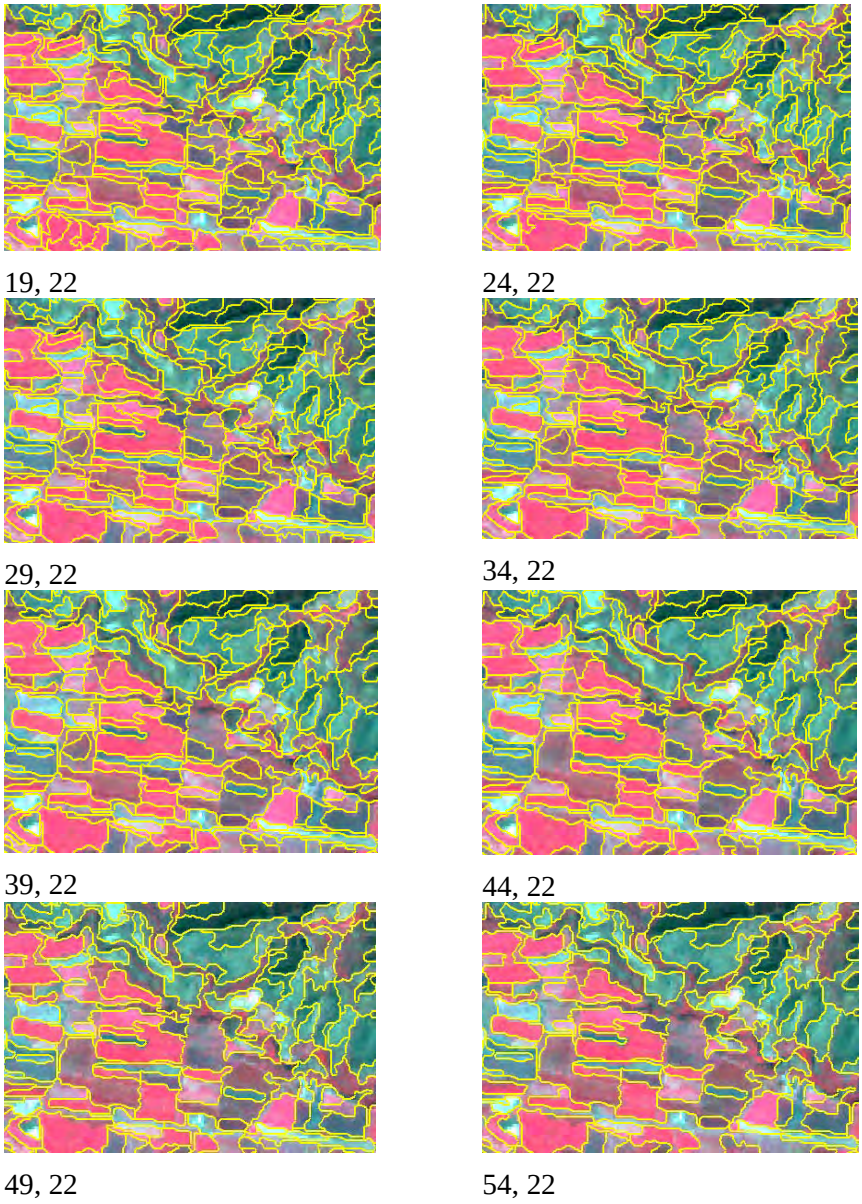
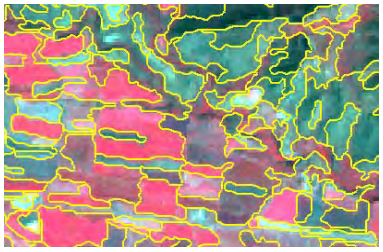


Figure 3-3: *Segmentations with "similarity" threshold from 19 to 59 with intervals of 5, and a constant "area" threshold of 22.*



59, 22

Figure 3-3 (continuation): *Segmentations with "similarity" threshold from 19 to 59 with intervals of 5, and a constant "area" threshold of 22.*

3.4.2. Evaluation of segmentations with objective function

3.4.2.1 Calculation of variance and spatial autocorrelation of segmented images.

For the nine segmentations, the variance and Moran's I were calculated and represented using the averaging values of those from each band (table 3-3). With the "similarity" threshold increasing from 19 to 59, the intrasegment variance increased from 325.684 to 559.811, and the intersegment spatial autocorrelation between neighbouring segments by Moran's I index decreased from 0.642 to 0.383. With a constant "area" threshold of 22, the larger the "similarity" threshold value the larger the generated segments, the higher the variance indexes, and the lower the intersegment autocorrelations are.

Table 3-3:
Variance (v) and Moran's I (I) indices of segmentations with experimented similarity and a constant area threshold (22).

Similarity	19	24	29	34	39	44	49	54	59
v	325.7	333.2	345.3	363.2	387.9	419.9	463.1	509.7	559.8
I	0.642	0.616	0.581	0.549	0.516	0.475	0.449	0.425	0.383

3.4.2.2 Evaluation of segmented images by the Objective Function.

The variance and autocorrelation indices of the nine segmentations were normalized, resulting in each of the indices whose values range between 0 and 1. The objective function for the segmentation of each band was calculated based on equation (1), and the obtained results are presented in figure 3-4.

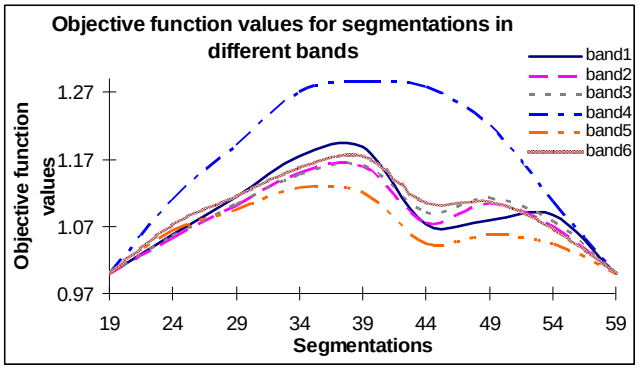


Figure 3-4: *Objective function values for segmentations with varying "similarity" threshold, and a constant "area" threshold from each band of Landsat image.*

The objective function values for bands 1, 2, 3, 5, and 6 present very similar distribution: from segmentation based on parameters (19, 22) the value starts to rise until it reaches its maximum at parameters (34, 22) and (39, 22), then the value started to go down until (44, 22) as the lowest then rise again and peaked at segmentation with parameters (49, 22), and (54, 22) then it goes down again. For band 4, first, the objective function values are higher than all those corresponding values from the other bands, and if deriving the objective function value by averaging the values from individual band, values from band 4 have bigger influence to the final result; second, it presents a bell-shaped distribution, it also peaks at segmentation (34, 22) (39, 22), and even (44, 22) then the value keeps going down until it reaches (59, 22). As a Near-Infrared band, band 4 is more sensitive to the changes of the spectral information from different vegetation types and allows better separability between classes of interest in this forest area, and it appears more sensitive to the segmentation parameter changes (table 3-4 and figure 3-4).

Table 3-4:
Normalized variance F(v) and Moran's I F(I) indices of segmentations by experimented "similarity" and a constant "area" threshold(22).

Similarity	19	24	29	34	39	44	49	54	59
F(v)	1.00	0.97	0.92	0.84	0.73	0.60	0.41	0.21	0.00
F(I)	0.00	0.10	0.21	0.33	0.45	0.52	0.70	0.86	1.00
F(v, I)	1.00	1.07	1.13	1.17	1.18	1.11	1.11	1.08	1.00

Figure 3-4 presented the calculated objective function values for each spectral band of Landsat image. The figure shows that except band 4 (near infrared band), the objective function of other spectral bands presents similar patterns and they are similar to the pattern presents by the accuracy assessment results. In order to observe the objective function of the entire multi-spectral image, two methods were used to "average" the objective function values of individual spectral bands. The first one is a simple average method, mean values were calculated and all the spectral bands were given the same weight and the result was presented in figure 3-5 the data series represented by pink line; the second way considered the possibly different influence of spectral bands in the contribution of the final result, based on the separability of the spectral bands to the classes of interest (figure3-6). Thus firstly the separability of the spectral bands to the classes of interest was calculated and presented in figure 3-6. The values were normalized into data range of 0-1, and the final objective function value for the entire multispectral image was calculated by adding those calculated weights to each bands. The final result was presented in figure 3-5.

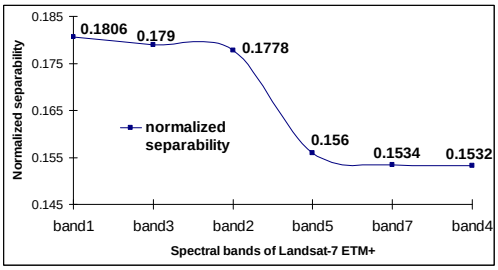


Figure 3-6: *Normalized separability for the spectral bands of Landsat image.*

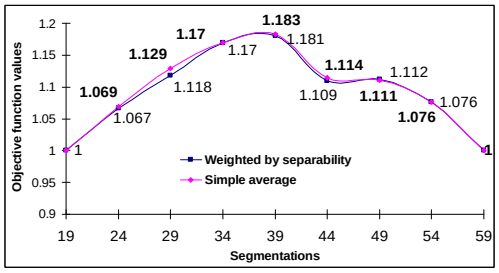


Figure 3-5: *Objective function values for the segmented images.*

Figure 3-5 shows that there is little, if any, difference between the objective

function values obtained using these two methods. The objective function values show that among the nine experimented segmentations, segmentations with parameters (34, 22), and (39, 22) obtained the highest values meaning that they are the best segmentations among all the tested ones.

To obtain the best segmentation result, the choice of parameters depends on the data type, the land-cover type, and which and how many spectral bands are used in image segmentation. Espindola *et al*. (2006) tested 2500 combinations of similarity (1-50) and area (1-50) on a single band of Landsat image in a small area to find the optimal combination of both parameters. In practice, multi-spectral data of large areas are typically used; thus for the selection of segmentation parameters visual inspection and objective function should be combined. Visual inspection can be used to rule out some results that are evidently over-or under-segmented. The objective function can then be used to determine the parameters for the best segmentation results.

3.4.3. Pixel and segment based classifications and accuracy assessment

To test the hypothesis that best segmentation results, according to the objective function, also yield the best classification results, the segmented images were classified by MLC. Eight land cover types are to be classified: irrigated agriculture, rain fed agriculture, temperate forest, orchards, grassland, tropical dry forest, lava flow, and human settlement. The classified images were evaluated by independent ground data which comprised of 600 randomly distributed points. The obtained accuracy assessment results were compared and listed in table 3-5 and figure 3-7.

Table 3-5:
Pixel-based and segment-based images classification accuracies (%).

Pixel	19,22	24,22	29,22	34,22	39,22	44,22	49,22	54,22	59,22
68.7	70.2	70.7	71	71.3	70.5	69.8	70.3	69.8	67.2

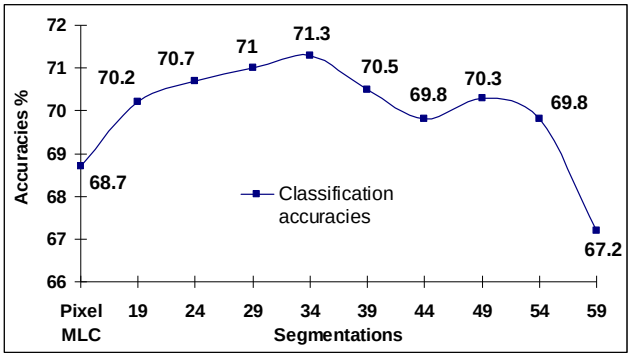


Figure 3-7: *Accuracies of pixel-based and segment based classifications, corresponding to table 3-5.*

The accuracy results presented very similar distribution as that of the objective function values. The accuracy starts to rise from segmentation (19, 22), until it reaches its highest value at segmentation (34, 22), the optimal segmentation with the highest objective function value, suggesting that the best segmentations indeed result in highest classification accuracies; the accuracy presents a low value at segmentation (44, 22) which also corresponds to a low objective function value. This close correspondence between the accuracy values and objective function values shows that the segmentation result has a direct influence to the classification result and the objective function seems effective to decide the optimal segmentation to carry out the object based classification. The segmentations on the left side were mostly over-, and on the right under-segmented. In the classification stage, the segments in the over-segmented images could still be allocated to their proper class, and more serious over-segmentation approximates pixel-based image classification. On the other hand, we do not know classification programs that can separate land-cover types which are grouped in one segment, i.e. deal with under-segmentation. Segmentation using (34, 22) obtained the highest object function value resulting in classification accuracy 4.1% higher than that obtained by the segmentation with the lowest objective function value. The test of McNemar (z^2 = 4.01) indicates that the difference between these two classification accuracies is significant with p= 0.05, p is the significance level, which shows that the classification based on the optimal segmentation is significantly better than that from the segmentation with a low objective function value, showing

that the objective function is indeed an effective way to determine the segmentations to carry out the classifications.

The entire experimented segment based classifications obtained higher accuracy than that by pixel based MLC, which showed that the segment based classification has advantage over the pixel based one. The classification based on the optimal segmentation (segmentation with the highest objective function value) and the pixel based MLC classification was compared. Based on table 3-1, among 600 ground data points, 384 points were classified correctly by both methods; 143 points were classified wrongly by both methods; 28 points were classified correctly by the pixel based MLC while incorrectly by the segment based method, and 45 points were classified correctly by the segment based method while incorrectly by the pixel based MLC. Based on equation 4, we calculated the chi-square statistic (z^2 = 3.96). Comparing the Chi-square distribution table we concluded that the segment based classification obtained accuracy that is significantly different from that obtained by pixel based MLC, with the significance level p = 0.05.

This study shows that the different levels of segmentation result in different accuracy values for land cover mapping. Segmentation compared to the one-pixel situation shows that segmentation indeed does provide better estimates, i.e., object based parameter estimation performs better than per-pixel estimation. By segmenting the images, information is lost. Up to a certain level, this is expected to be noise stemming either from spectral noise or spatial mismatch. Spectral noise can stem from sensor noise, atmospheric influences, or the effects of the point-spread function (Schowengerdt 1997). At a certain aggregation level, the lost information might turn out to be relevant. This would show in worse results, in our case the misclassifications. Both situations can be observed in the accuracy curve for the different scale parameters. This study does not aim at determining the exact scale parameters that yield optimal segmentation and then the highest classification accuracy, but merely at showing that different heterogeneity values yield different results. The optimal scale of observation (i.e., object size with object based image classification) depends on: (a) the scale of the phenomenon of interest, the process which are responsible for it and how they can be scaled up, and (b) on the spatial

heterogeneity of the landscape. The setting of the "area" parameter for the segmentation was held constant in this study by using "22" for all segmentations. However, different "area" parameters produce segmentation with different spatial variances. A combination of different "similarity" and "area" threshold was not considered but will be a future research topic.

3.5. Conclusions

Image segmentation is a preliminary and critical step in OBIA. Evaluating and optimizing segmentation results prior to proceeding to any further analysis ensures the most suitable input for the image classification. In this paper, a test study using medium-resolution Landsat ETM+ image was carried out. The segmentation package SPRING, previously identified as the best non-commercial segmentation program, was chosen for the image segmentation. Nine segmentations with "similarity" thresholds ranging from 19 to 59 in intervals of 5, and an "area" threshold of 22 were performed, and the results evaluated with an objective function (Espindola *et al*. 2006). Results show that intrasegment variance increased and intersegment autocorrelation decreased with an increasing segment size. The evaluation results by the objective function for the nine segmentations peaked at (34, 22) and (39, 22), which show a good balance of intrasegment variance and intersegment correlation. The segmented images were subsequently classified with MLC, and the results evaluated. The best segmentations, in terms of objective function rating, also led to the most accurate classification, and, the accuracy values presented similar distribution as the objective function values in the function of the segmentations. Results also showed that although neither over-nor under-segmentation are desirable and show low objective function values, in the classification stage, over-segmented images obtained better results than under-segmented images. Although neither over- or under-segmentations are desirable, comparatively, over-segmentation is a less significant problem in the resulting classification, as the segments can still be allocated to their accurate land-cover types. However, under-segmentation is a more serious problem since we are not aware of any programs that are able to separate different land-cover types from one segment. The test of McNemar showed that classification based on the optimal segmentation is significantly better than from the

segmentation with the lowest objective function value. This result showed that the objective function is indeed an effective way to evaluate the segmentation result in order to select the optimal image segmentation for the region based image classification. Classification based on the optimal segmentation is significantly different from that of the pixel based MLC result, which showed that the OBIA has advantage over the pixel based classification and is able to obtain higher classification accuracy.

Acknowledgements

The authors thank CONACYT-CONAFOR 2005-C02-14741 for supplying PhD fellowship to the first author during writing up of the paper.

References

Adams, R., and Bischof, L., 1994, Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16, 641-647.

Arvidson, T., J. Gasch & S.N. Goward, 2001, Landsat 7's long term acquisition plan – an innovative approach to building a global imagery archive. Remote Sensing of the Environment, 78, No. 1, pp. 13-26.

Benz, U. C., Hofmann P., Willhauck G., Lingenfelder I., Heynen M., 2004, Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS Journal of Photogrammetry & Remote Sensing, 58, 239-258.

Bins, L. S. A., Fonseca, L.M., Erthal, G.J., Ii, F.M., Satellite Image Segmentation: a region growing approach. VIII Simposio Brasilero de Sensoramiento Remoto, Salvador, Bahía, Brasil. 14-19 April 1996.

Blaschke, T., Land, S., Lorup, E., Strobl, J., and Zeil, P., 2000, Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications. Environmental Information for Planning, Politics and the Public, Metropolis Verlag, Marburg, Volume II.

Blaschke, T., Strobl, J., 2001, What's wrong with pixels? Some recent development interfacing remote sensing and GIS. Interfacing Remote Sensing and GIS 6/01, 12-17.

Câmara, G., Souza R. C. M., Freitas U. M., Garrido J., 1996, SPRING: Integrating remote sensing and GIS by object-oriented data modelling. Computers & Graphics, 20: (3)395-403.

Definiens, 2006. Definiens professional User Guide 5. Definiens AG, Munich.

Ebert, A., 2006, Social vulnerability assessment using satellite data. A case study for Tegucigalpa. Friedrich-Schiller-Universitat Jena Chemisch-Geowissenschaftliche Fakultat Institut fur Geographie & Int. Institute for Geo-Information Science and Earth Observation, Department of Earth Observation Science Enschede, the Netherlands, Master thesis, pp. 50-52.

Emrahoglu, N., Yegingil, I., Pestemalci, V., Senkal, O., Kandirmaz, H.M., 2003, International Journal of Remote Sensing, 24, 649-655.

Espindola, G. M., Camara, G., Reis, I. A., Bins, L. S., and Monteiro, A. M., 2006, Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. International Journal of Remote Sensing, 27, 3035-3040.

Evans, C., Jones, R., Svalbe, I., and Berman, M., 2002, Segmenting multispectral Landsat TM images into field units. IEEE Transactions on Geoscience and Remote Sensing, 40, 1054-1064.

Egawa, H., and Kusaka, T., 1988, Region extraction in SPOT data. *Geocarto International* 3, 25-32.

Flanders, D., Hall-Beyer, M., and Perenerzoff, J., 2003, Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. *Canadian Journal of Remote Sensing*, 29, 441-452.

Flynn, L.P., A.J.L. Haris & R. Wright, 2001, Improved identification of volcanic features using Landsat-7 ETM+. Remote Sensing of Environment, 78, No. 2, pp. 180-193.

Foody, G., 1999, Image classification with a neural network: from completely-crisp to fully fuzzy situations. In: Atkinso/Tale (eds.), Advances in Remote Sensing and GIS analysis. Wiley & Son, Chichester, 17-37.

Foody, G.M., 2004, Thematic Map Comparison: Evaluating the Statistical Significance of differences in Classification Accuracy. Photogrammetric Engineering & Remote Sensing, 70, 627-633.

Foody, G.M., and Mathur, A., 2006, The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed

spectral responses for classification by a SVM. Remote Sensing of Environment, 103, 179-189.

Foody, G.M., Mathur, A., Sanchez-Hernandez, C., and Boyd, D.S., 2006, Training set size requirements for the classification of a specific class. Remote Sensing of Environment, 104, 1-14.

Giada, S., De Groeve, T., Ehrlich, D., 2003, Information extraction from very high resolution satellite imagery over Lukole refugee camp, Tanzania. International Journal of Remote Sensing, 24, 4251-4266.

Gitas, I.Z., Mitri, G.H., Ventura, G., 2004, Object-based image classification for burned area mapping of Creus Cape, Spain, using NOAA-AVHRR imagery. Remote Sensing of Environment, 92, 409-413.

Gorte, B., 1998, Probabilistic segmentation of remotely sensed images. ITC Publication Series No. 63.

Goward, S.N., J.G. Masek, D.L. Williams, J.R. Irons & R.J. Thompson, 2001, The Landsat 7 mission. Terrestrial research and applications for the 21st century. In: Remote Sensing of Environment. 78, No. 1, pp. 3-12.

Haralick, R. M., and Shapiro, L. G., 1985, Image segmentation techniques. Computer Vision Graphics and Image Processing, 29, 100-132.

Ho, T.K., and Basu, M., 2002, Complexity measures of supervised classification problems. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 289-300.

Huang, C., Davis, L.S., and Townshend, J.R.G., 2002, An assessment of support vector machines for land cover classification. International Journal of Remote Sensing, 23, 725-749.

Ichoku, C., and Karnielli, A., 1996, A review of mixture modeling techniques for sub-pixel land cover estimation. Remote Sensing Reviews, 13, 161-186.

Jensen, J.R., 1996, Introductory digital image processing: A remote sensing perspective, 2nd Ed. New Jersey: Prentice Hall Inc. 379 pp.

Lang, S., Blaschke, T., 2006, Bridging Remote Sensing and GIS-What are the main supportive pillars? 1st International Conference on Object-based Image Analysis (OBIA 2006).

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L., *et al.* 2000, Development of a global land cover characteristics database and IGBP DISCover from 1km AVHRR data. International Journal of

Remote Sensing, 21, 1303-1330.

Mas, J-F., and Ramirez, I., 1996, Comparison of land use classifications obtained by visual interpretation and digital processing. International Journal of Applied Earth Observation and Geoinformation, 1996-3/4, 278-283.

Mather, P., 2004, Computer processing of remotely-sensed images: an introduction, 3rd Ed. San Francisco, CA: John Wiley and Sons Ltd. 324pp.

Meinel, G. & Neubert, M. (2004): A Comparision of segmentation programs for high resolution remote sensing data. Proceedings XXth ISPRS Congress, Istanbul, 14.-23. July 2004, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV-B4, p. 1097-1102 (ISSN 1682-1750).

Milton, E., 1999, Image endmembers and scene models. In Canadian Journal of Remote Sensing 25(2): 112-120.

Sader, S.A., and Winne, J.C., 1992, RGB-NDVI colour composites for visualizing forest change dynamics. International Journal of Remote Sensing, 13, pp. 3055-3067.

Schowengerdt , R., 1997, Remote sensing models and methods for image processing. Academic press, San Diego.

Skidmore, A., Turner, B., Brinkhof, W., Knowles, E., 1997, Performance of a Neural Network: mapping forests using GIS and remotely sensed data. Photogrammetric Engineering & Remote Sensing 63, 501-514.

Stuckens, J., Coppin, P. R., and Bauer, M. E., 2000, Integrating contextual information with per-pixel classification for improved land cover classification. Remote Sensing Environment, 71, 282-298.

Townshend, J., Huang, C., Kalluri, S., Defries, R., Liang, S., Yang, K., 2000, Beware of per-pixel characterisation of land cover. International Journal of Remote Sensing, 21 (4): 839-843.

CHAPTER 4

A COMPARISON OF THE PERFORMANCE OF PIXEL-BASED AND OBJECT-BASED CLASSIFICATIONS OVER IMAGES WITH VARIOUS SPATIAL RESOLUTIONS[1]

Abstract: In the last two decades, advances in computer technology, earth observation sensors and GIS science, led to the development of "Object-based Image Analysis (OBIA)" as an alternative to the traditional pixel-based image analysis method. It was recognized that traditional pixel-based image analysis

[1] Based on the contents of this chapter, the article "A comparison of the performance of pixel based and object based classifications over images with various spatial resolutions" with the authors: Yan Gao and Jean F. Mas is published in Online Journal of Earth Sciences, in 2008, volume 2 (1), pp. 27-35. ISSN: 1991-7708.

Another article "Comparison of pixel-based and object-oriented image classification approaches - a case study in a coal fire area, Wuda, Inner Mongolia, China." With authors Yan Gao, Mas, J. F., Maathuis, B. H. P., Zhang, X. M., and Van Dijk, P. M., is published in International Journal of Remote Sensing, in 2006, Volume 27, pp. 4039-4051.

Based on the content of this chapter, a paper titled "A comparison of the performance of pixel-based and object-based classifications over images with various spatial resolutions" was presented in the conference GEOBIA 2008, 6-7, August, 2008, Calgary, Canada.

is limited because of the following reasons: image pixels are not true geographical objects and the pixel topology is limited; pixel based image analysis largely neglects the spatial photo-interpretive elements such as texture, context, and shape; the increased variability implicit within high spatial resolution imagery confuses traditional pixel-based classifiers resulting in lower classification accuracies (Hay and Castilla 2006). Different from pixel-based method, OBIA works on (homogeneous) objects produced by image segmentation and more elements can be used in the classification. As an object is a group of pixels, object characteristics such as mean value, standard deviation, ratio, etc can be calculated; besides there are shape and texture features of the objects available which can be used to differentiate land cover classes with similar spectral information. These extra types of information give OBIA the potential to produce land cover thematic maps with higher accuracies than those produced by traditional pixel-based method. In this chapter, the performance of OBIA with different spatial resolution satellite images was studied. By comparing the object based classification result with that produced by the pixel based one, this work intends to find out how spatial resolution of satellite images influences the performance of OBIA. Experiment results showed that with the two sets of images of four different spatial resolutions, object based image analysis obtained higher accuracies than those of the pixel based ones; with the increasing of the spatial resolution, the difference in accuracy values between object based and pixel based is decreasing. This paper showed that the object-based image analysis has advantage over the pixel-based one, and in accuracy rating, the advantage was better represented by higher spatial resolution satellite images.

Key words: Pixel-based image analysis; Object-based image analysis; accuracy assessment; simulated images.

4.1. Introduction

There is a steadily increasing need for timely and accurate geo-spatial information. The automatic classification of remotely sensed data is an essential action within the process of generating or updating GIS databases. Remote sensing image classification is a commonly adopted method to obtain

land cover information from Satellite images and many classification algorithms have been extensively applied to (Ouattara 2004, Borak *et al*. 1999, Chintan *et al*. 2004, Casals-Carrasco *et al*. 2000). Since remote sensing images consist of rows and columns of pixels, naturally conventional land cover mapping has been based on single pixels (Dean *et al*. 2003). Pixel based image classification utilizes spectral information-digital values (DNs) stored in the image and classifies images by considering the spectral similarities with the pre-defined land cover classes (Casals-Carrasco *et al*. 2000). Although the techniques are well developed and have sophisticated variations such as soft classifiers, sub-pixel classifiers and spectral un-mixing techniques, it is argued that it does not make use of the spatial concept (Blaschke *et al*. 2000). For example Zhou (2001), Pizzolato *et al*. (2003), Dean (2003) claimed that Maximum likelihood classifier (MLC) was limited by utilizing spectral information only without considering contexture information. And texture information was ultimately necessary if one is to obtain accurate image classifications (Zhou 2001).

The concept of OBIA as an alternative to pixel based analysis was introduced in 1970s (de Kok *et al*. 1999). The initial practical application was towards automation of linear feature extraction (Flanders *et al*. 2003). In addition to the limitation from hardware, software, poor resolution of images and interpretation theories, the early application of object based image analysis faced obstacles in information fusing, classification validation, reasonable efficiency attaining, and analysis automation (Flanders *et al*. 2003). Since mid-1990s, hardware capability has increased dramatically and high spatial resolution images with increased spectral variability became available. Pixel based image classification encountered serious problem in dealing with high spatial resolution images and thus the demand for OBIA has increased (de Kok *et al*. 1999). OBIA works on objects instead of single pixels. The idea to classify objects stems from the fact that most image data exhibit characteristic texture features which are neglected in conventional classifications (Blaschke *et al*. 2001). In the early development stage of OBIA, objects were extracted from pre-defined boundaries, and the following classifications based on those extracted objects exhibited results with higher accuracy, comparing with those by pixel based methods (Smith *et al*. 2001, Dean *et al*. 2003, de Wit *et al*.

2004). This technique - classifying objects extracted from predefined boundaries - is applicable for agriculture plots or other land cover classes with clear boundaries, while it is not suitable to the areas with no boundaries readily available, such as semi-natural areas. Image segmentation is the solution for obtaining objects in areas without pre-defined boundaries. It is a preliminary step in object-based image analysis. By image segmentation, image is divided into homogeneous, continuous, and contiguous objects. Image segmentation is a kind of regionalization which is to divide the image regarding to a certain criteria of homogeneity, and at the same time, requires spatial contingency. There are many techniques to perform image segmentation and those techniques can be categorized into three classes: thresholding/clustering, region based, and edge based (Fu and Mui 1981, Haralick and Shapiro 1985). Image segmentation is analogous to the common multi-spectral clustering of individual pixels but with grouping taking place in the spatial domain in addition to the multispectral domain (Lobo 1997). A more detailed description and evaluation of segmentation is in methodology part. Based on image segmentation, OBIA uses information from spectral, textural and contextual, and spatial domain. Particularly, image objects allow shape characteristics and neighbourhood relationships to be used for the object's classification.

Due to the problems encountered in thematic mapping, much work has been done to increase the accuracy of thematic maps which are derived from remote sensing data (Foody 2004). Thus, much work has been done to compare different classification techniques for example between pixel based and object based classifications (Dean 2003), and in those studies, the key focus is on the difference among the estimated classification accuracies.

This paper looks at the performance of OBIA with different spatial resolution satellite images; comparing the classification results with those produced by the pixel-based method, it intends to find out how spatial resolution of satellite images influences the performance of OBIA. Images with four spatial resolutions were used: 10m, 30m, 100m, and 250m; the coarse spatial resolution images were generated from a SPOT-5 image with 10 m spatial resolution, one set by mean filtering, and another using a method simulating the real response of a satellite sensor moving across the scene, which was based on

the algorithm designed by Justice *et al*. (1989).

4.2. The study area and Data

4.2.1 The study area

It is located at south-east of Tancítaro mountain peak. It covers approximately 27*28 km^2, ranging in latitude from approximately 19º 02' to 19º 17' N and in longitude 102º 00' to 102º 16'W (figure 4-1). The dominant land cover types are: irrigated agriculture, grassland, tropical dry forest, human settlement, orchards, and temperate forest.

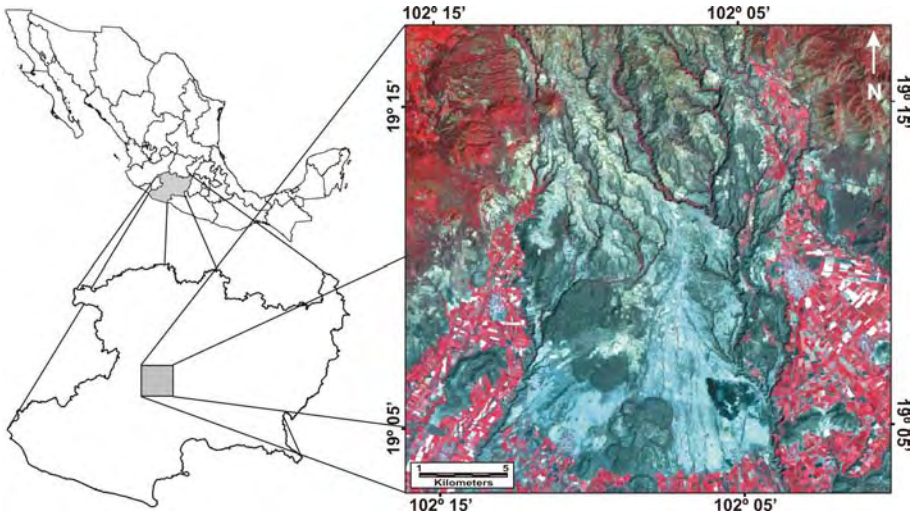


Figure 4-1: *The study area. The satellite image was presented with false colour composite of RGB bands with near infrared, red, and green.*

4.2.2 Data: There are three types of images used in this research: SPOT-5, obtained on 2004, 13, March; Landsat-7 Enhanced Thematic Mapper plus (ETM+) image obtained on 16 February 2003; and Moderate Resolution Imaging Spectro-radiometer (MODIS) image obtained March 2005.

The orbital type of SPOT-5 is sun-synchronous, with orbital altitude of 822 km. The sensor has a swatch width of 60 km and the acquired data is 8 bit per pixel.

The panchromatic sensor takes the image in one channel within spectral range of 0.51-0.73 μm and it has spatial resolution of 5m / 2.5m. The characteristics of multi-spectral sensor of SPOT-5 are presented in table 4-1.

Table 4-1:
The characteristics of SPOT-5 Imagery.

Spectrum bands	Width (μm)	Spatial resolution	Depth (bit)
1 visible (green)	0.50 – 0.59	10 m	8
2 visible (red)	0.61 – 0.68	10 m	8
3 visible (NIR)	0.78 – 0.89	10 m	8
4 near infrared (NIR)	1.58 – 1.75	20 m	8

Table 4-2:
The characteristics of Landsat-7 ETM Imagery.

Spectrum bands	Width (μm)	Spatial resolution	Depth (bit)
1 visible (blue)	0.45 – 0.51	30 m	8
2 visible (green)	0.52 – 0.60	30 m	8
3 visible (red)	0.63 – 0.69	30 m	8
4 near infrared (NIR)	0.75 – 0.90	30 m	8
5 middle Infrared (NIR2)	1.55 – 1.75	30 m	8
6 thermal Infrared (TIR)	10.40 – 12.50	30 m	8
7 middle Infrared (MIR)	2.09 – 2.35	30 m	8
8 panchromatic (PAN)	0.52 – 0.90	15 m	8

(MODIS) is a key instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites. Terra's orbit around the Earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral channels. These data can be used to research global dynamics and processes occurring on the land, in the oceans, and in the lower atmosphere. The altitude of its operating orbit is 705km, 10:30 am is the descending node for Terra and 1:30 pm is ascending node for Agua. The orbit is sun synchronous, near polar

circular. The instrument scan the data in a cross track way and the swath dimensions are 2330km across track and 10km along track at nadir. The image data recorded by MODIS is 12 bits per pixel. The spatial resolution ranges from 250m to 500m and to 1000m. The designed life for MODIS is 6 years. For mapping land use and land cover, MODIS spectral bands 1-7 are closely related. The characteristics of the first seven bands of MODIS are presented in table 4-3.

Table 4-3:
The characteristics of MODIS imagery (the first seven bands).

Spectrum bands	Width (μm)	Spatial resolution	Depth (bit)
1 visible (red)	0.62 – 0.67	250 m	12
2 visible (NIR)	0.841 – 0.876	250 m	12
3 visible (blue)	0.459 – 0.479	500 m	12
4 near infrared (green)	0.545 – 0.565	500 m	12
5 middle Infrared (MIR)	1.23 – 1.25	500 m	12
6 thermal Infrared (MIR)	1.628 – 1.652	500 m	12
7 middle Infrared (MIR)	2.105 – 2.155	500 m	12

Spectral bands 8-36 are not closely related to this research, and their primary use is summarized relatively simple. Bands 8-16 are mainly used in Ocean Colour/Phytoplankton/Biogeochemistry monitoring; bands 17-19 are mainly used in Atmospheric/Water Vapor mapping; bands 20-23 are mainly used in Surface/Cloud temperature mapping; bands 24-25 are used for mapping atmospheric temperature; bands 26-28 are used for mapping Cirrus Clouds and water vapor monitoring; band 29 is used for mapping cloud properties and band 30 is used for mapping ozone; bands 31-32 are used for mapping surface/cloud temperature and bands 33-36 are used to mapping cloud top altitude (http://modis.gsfc.nasa.gov/data/dataprod/index.php, accessed on 22 January 2008). Pixel based image classification was carried out in ILWIS (ILWIS 2005), and OBIA was carried out in eCognition (Definiens, 2006).

4.3. Methods

4.3.1. Image pre-processing

The SPOT-5 image was corrected geometrically: 28 evenly distributed Ground Control Points (GCPs) were collected from ortho-corrected photographs, which have spatial resolution of 2m. SPOT-5 image was registered to the UTM Zone 13, WGS84 coordinate system. After the correction, image fit was considered acceptable and the RMS error was smaller than one pixel size.

Landsat-7 ETM+ image was geometrically corrected by 86 evenly distributed ground control points (GCPs) extracted from a mosaic of ortho-photographs. The rectification was carried out in a sub-pixel level and the ERM error was less than a pixel (16.8m). The image was re-sampled by Nearest Neighbour and was geo-referenced to UTM zone 13, with the spheroid and datum in WGS 84.

4.3.2. Generating the multi-spectral images with coarser spatial resolutions

The original four bands of SPOT-5 imagery were filtered into a range of resolutions, using two different methods. The first method was a simple mean filter which assigned the mean reflectance value of a user-defined window of pixels to a single pixel covering the whole area of the window. The second method attempted to simulate more accurately the real response of a satellite sensor moving across the scene, which was based on the algorithm designed by Justice et al. (1989) to approximate Multi-Spectral Scanner (MSS) sampling. Justice *et al*. (1989) used this filter to achieve a range of resolutions by successively filtering an image, adding noise based on the original variance of the data, and then re-filtering with the same kernel. However, for this study, it was decided to interpolate between the original kernel values to generate kernels ranging from 3*3 to 25*25, so that a single filtering operation would be sufficient to produce each coarser image. As with the mean filter, this filter calculated one value for each window of cells, and created a coarser cell containing this value (Bastin 1997).

The mean filter tended to smooth the image, while cubic filtering retained differences between the resulting pixels, in other words, retained the local

contrast (figure 4-2). When local image variance was calculated for these filtered images it was higher at any point in the cubic filtered image than in the corresponding mean filtered image. This effect is particularly relevant to analysis which searches for edges and patterns in an image, such as the use of split moving windows (Cornelius and Reynolds 1991) to look for discontinuities in a scene's reflectance values. Figure 4-2 showed that by degrading SPOT imagery, images with coarser spatial resolutions were produced. With the increasing of the pixel size, more and more mixed pixels-pixels with spectral information coming from different land cover types-were generated. In the simulated images with pixel size of 250m, the images are composed mainly of mixed pixels.

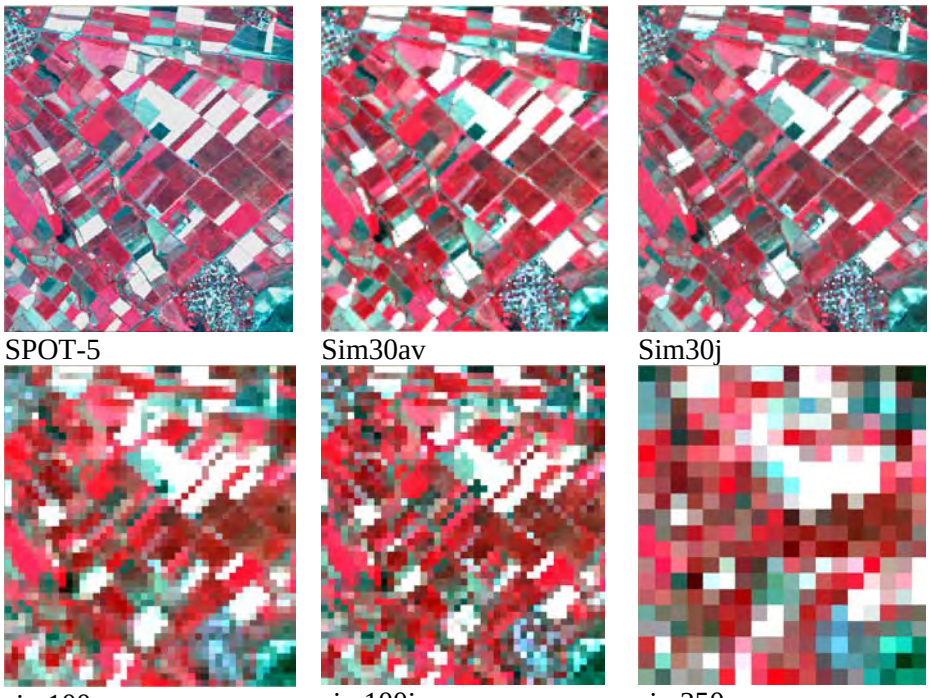


Figure 4-2: *The simulated images by mean-filtering (av) and by cubic-filtering (j). For example, sim100av means simulated image generated by mean filtering with 100 m spatial resolution; sim100j means simulated image generated by cubic-filtering with 100m spatial resolution.*

sim250j

Figure 4-2 (continuation): *The simulated images by mean-filtering (av) and by cubic-filtering (j). For example, sim100av means simulated image generated by mean filtering with 100 m spatial resolution; sim100j means simulated image generated by cubic-filtering with 100m spatial resolution.*

4.3.3. Pixel-based image classification

Classical pixel-based image classification automatically categorizes all pixels in an image into land cover classes or themes in a pixel by pixel manner. Usually, multispectral data are used to perform the classification and the spectral pattern present within the data for each pixel is used as the numerical basis for categorization. That is, different feature types manifest different combinations of digital numbers based on their inherent spectral reflectance and emittance properties. In this point of view, a spectral "pattern" is not geometric in character but refers to the set of radiance measurements obtained in the various wavelength bands for each pixel. The classical pixel-based methods are minimum-distance/nearest neighbour, parallelepiped and maximum likelihood classifiers (MLC), whose detailed information can be found in Lillesand and Kiefer (1994). In land cover classification through the satellite sensor data, pixel-based supervised MLC is the most popular statistical algorithm and is widely accepted as a standard approach (Emrahoglu *et al.* 2003). It allocates a pixel to the class with which it has the highest probability of membership where the likelihood $L_i(x)$ that a pixel x is a member of class i is given by equation 1:

$$L_i(x) = (2\pi)^{-n/2}\left|V_i\right|^{-1/2} e^{-y/2} \qquad (4.1)$$

in which V_i is the covariance matrix of class i, n is the number of spectral bands, and y is the Mahalanobis distance. The rescaling of $L_i(x)$ between 0 and 1 yields the MLC a posteriori probability $P_i(x)$ (Dean *et al*. 2003).

4.3.4. Object-based image analysis

OBIA does not operate directly on single pixels, but on objects consisting of many pixels, that have been grouped together in a meaningful way by image segmentation. The accuracy of segmentation directly affects the performance of OBIA. Recently, Espindola *et al.* (2006) proposed an objective function to evaluate the segmentation results and experiment results showed that it is an effective method to decide the optimal segmentation for the following OBIA (see chapter 3). In this chapter, objective function is not used to evaluate the segmentation as the classification is based on a multi-resolution segmentation by which the same image was segmented into objects in different scales and represented in different levels.

OBIA in eCognition comprises two parts: 1) (multi-resolution) image segmentation and 2) classification based on objects' features in spectral and spatial domains. Image segmentation is a kind of regionalization, which delineates objects according to a certain homogeneity criteria and at the same time requiring spatial contingency (Lang *et al*. 2006). Multi-resolution segmentation segments image objects in multiple scales. Ideally, the generated contiguous image segments represent real world objects on different levels of scale. In Definiens Professional 5, the created segments are embedded into a hierarchical network in which each object knows its neighbouring objects in horizontal and vertical direction (Baatz and Schape 1999, Hofmann 2001a). The structure of multi-resolution segmentation is shown in figure 4-3.

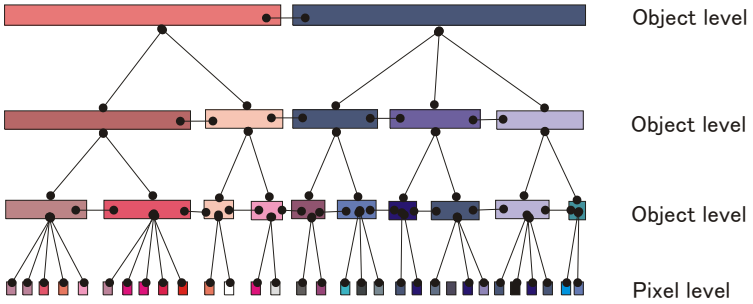


Figure 4-3: *Structure of multi-resolution segmentation. Objects generated from different scales have topological relations between them, so each object knows its super- and sub- and neighbouring objects. This topological relationship allows objects to be defined not only with spectral, shape, texture information, but with this semantic information (relations) between objects.*

The lowest level in the image object is the pixel level. After defining various parameters, such as the relative average segment size, compactness, smoothness and the influence of spectral properties, small segments are created in the second image level during a segmentation run. Once the first image level with small image objects is created, the objects are further aggregated in higher image levels. The original segment outlines remain but are aggregated in higher image levels. The way of fusing the small segments to bigger segments depends on the choice of segmentation parameters (average segment size, spectral and shape properties) that have to be specified individually for every image level.

There are several parameters to guide the segmentation result. Scale parameter is an abstract term which determines the maximum allowed heterogeneity for the resulting image objects. Colour criterion defines the weight the spectral values of the image layers contribute to image segmentation, as opposed to the weight of the shape criterion. There is the relation between colour and shape criteria: colour + shape = 1. Maximum colour criterion 1.0 results in objects spectrally most homogeneous; while with a value less than 0.1, the created objects would not be related to the spectral information at all. Smoothness is to optimize image objects with regard to smooth borders, and compactness is to optimize image objects with regard to compactness (Baatz *et al.* 2004). The

resulting objects also depends on the image data, for a given set of segmentation parameters, heterogeneous image data result in smaller image objects than homogeneous image data.

The potential of the multi-resolution segmentation is that, depending on the image resolution, the smallest objects that can be identified in the image can be classified using small image objects generated using a small scale value. Larger homogeneous areas can be classified using larger image objects on a higher image level with a bigger scale value. A hierarchical network of image objects is constructed where the resulting objects know their sub- super- and neighbour-objects. Based on this, relations between different classes in different image levels can be used as a criterion for the class description.

Another advantage of the multi-resolution segmentation is the introduction of image levels with thematic information that can be referred to. That is similar to the overlay algorithms as used in a Geographic Information System (GIS). For this study for instance, the settlement has been digitized manually and its existence has been included in the image analysis as thematic information. To avoid confusion and to reduce the error, the implementation of additional knowledge yields more accurate results.

Once the segmentation is completed, the created image objects can then be described and classified by an extensive variety of features that include colour-, texture-, form-, and context properties in several forms. This can be done using two classifiers, a (standard) nearest neighbour classifier (NN) and fuzzy membership functions, or a combination of both. The first describes the class by user-defined sample objects, while the second classifier describes intervals of feature characteristics (Hofmann 2001b). The variety of object features can be used either to describe fuzzy membership functions, or to determine the feature space for NN. There are many object features available including spectral related such as mean layer values, standard deviation, ratio, etc, and spatial related information such as objects' size, shape, texture. More importantly, there is semantic information available between objects in the same level and objects of its super- and sub-objects in one or more than one upper or lower levels. To fully apply the available information, the knowledge

to the study area or the interested objects is critical. More detailed description of image segmentation and classification can be found in Definiens Professional 5, Hofmann (2001a), and Gao *et al*. (2006).

4.3.5. Accuracy assessment of the classification results

Accuracy assessment quantifies data quality so that map users may evaluate the utility of a thematic map for their intended applications (Stehman and Czaplewski 1998). The land cover classifications from the map are compared to the reference classifications, and the extent to which these two classifications agree is defined as map accuracy. To perform a statistically rigorous accuracy assessment, it is required a probability sampling design and statistically consistent estimators of accuracy parameters, along with a response design determined in accordance with features of the mapping and classification process such as the land-cover classification scheme, minimum mapping unit, and spatial scale of the mapping (Stehman and Czaplewski 1998). The reference or "true" classification is obtained for each sampling unit based on interpreting aerial photography, field survey, the opinion of experts, or a combination of these sources. This process is called reference design. The reference classification is composed of the sampling units which provide the link between a spatial location on the map and the corresponding spatial location on the Earth. There are two types of sampling units: points or area units. Pixels and polygons are examples of area sampling units that are directly associated with mapped land cover features. Janssen and van der Wel (1994 p. 422) and Franklin *et al*. (1991) advocate using pixels as the basis of an accuracy assessment, the former arguing that "remote sensing data should be considered to be point-sampled data, in which the points possess a certain spatial extent". Janssen and van der Wel (1994) further state that individual pixels are the most appropriate sampling unit for a pixel-based classification. Restricting the assessment to homogeneous areas can avoid the confounding of classification and location error but it is not a recommended strategy because of the optimistic accuracy results that typically arise (Hammond and Verbyla 1996).

Error matrix is a two-dimensional matrix formed by land cover classification

and reference classification/ground data. It effectively summarizes the key information obtained from the sampling and response designs and it has been widely used in the image classification accuracy assessment (Smits *et al*. 1999, Tarmo *et al*. 2002, Foody 2004). The error matrix represents a contingency table in which the diagonal entries represent correct classifications, or agreement between the map and reference data, and the off-diagonal entries represent misclassifications, or lack of agreement between the map and reference data. Various summary measures are derived from the error matrix to describe accuracy. A core set of accuracy parameters that can be interpreted as probabilities defined for the map being assessed includes the following: *overall accuracy* describes the overall proportion of area correctly classified which represents the probability that a randomly selected point location is classified correctly by the map, *user's accuracy* describes the conditional probability that a randomly selected point classified as category I by the map is classified as category I by the reference data, *producer's accuracy* describes the conditional probability that a randomly selected point classified as category j by the reference data is classified as category j by the map.

Omission Error and Commission Error are also common in evaluating the accuracy of a thematic map. The Commission Error can be seen as the probability that, given a generic pixel of the thematic product, it does not belong to the same class in the reference data, and the Omission Error is the probability that, given a pixel of the reference data, it has not been correctly classified in the thematic product. Commission and Omission Error are linked to User's Accuracy (Ua) and Producer's Accuracy by the simple relationship: Commission Error = 1 - User's accuracy; Omission Error = 1 - Producer's accuracy (Boschetti *et al*. 2004).

Both pixel-based and object-based image analysis results were evaluated with independent reference data which were generated by interpreting in total 420 points generated with a stratified random sampling method. Based on the land use map, 60 random points were put in each of the 7 classes of interest. The properties of these points were interpreted based on the information from photographs, land use map from year 2000, and satellite images. In the study, classifications were performed to images with four different types of

resolutions: 10m, 30m, 100m, and 250m. With the increasing of the spatial resolution, reference data for images with 10m spatial resolution may not represent the property of images with 100m and 250m spatial resolution. In order to evaluate the classifications of the images with different spatial resolution, the same number of reference data in the same location was interpreted for images with 10m, 30m, 100m, 250m spatial resolutions, also for images of Landsat-7 ETM+ and MODIS. So altogether there are six sets of reference data to evaluate the accuracy of the classifications.

4.4. Results and discussion

4.4.1 Spectral separability analysis of land-cover classes

Figure 4-4 shows the scatter plots of SPOT-5 image four spectral band combinations of green, red, near infrared, and mid-infrared.

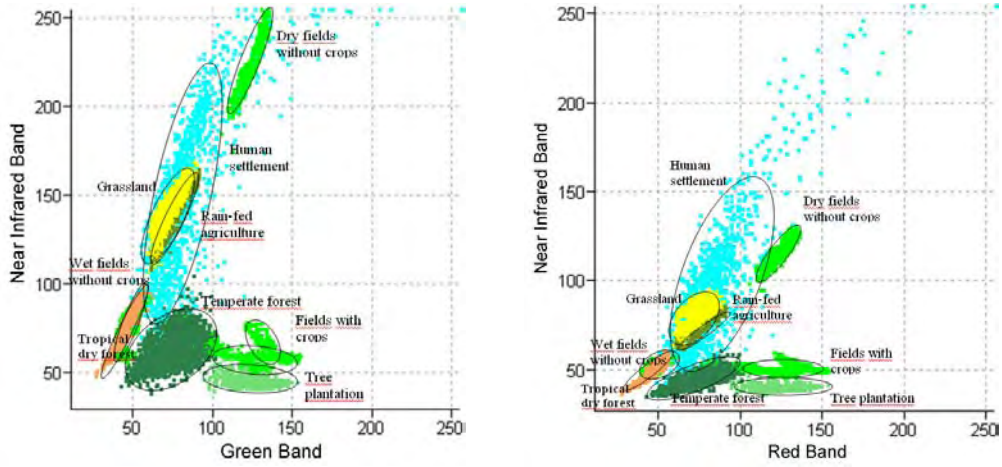


Figure 4-4: *Feature space for spectral bands combinations of green, near infrared and red, near infrared, in the example of SPOT-5 image*

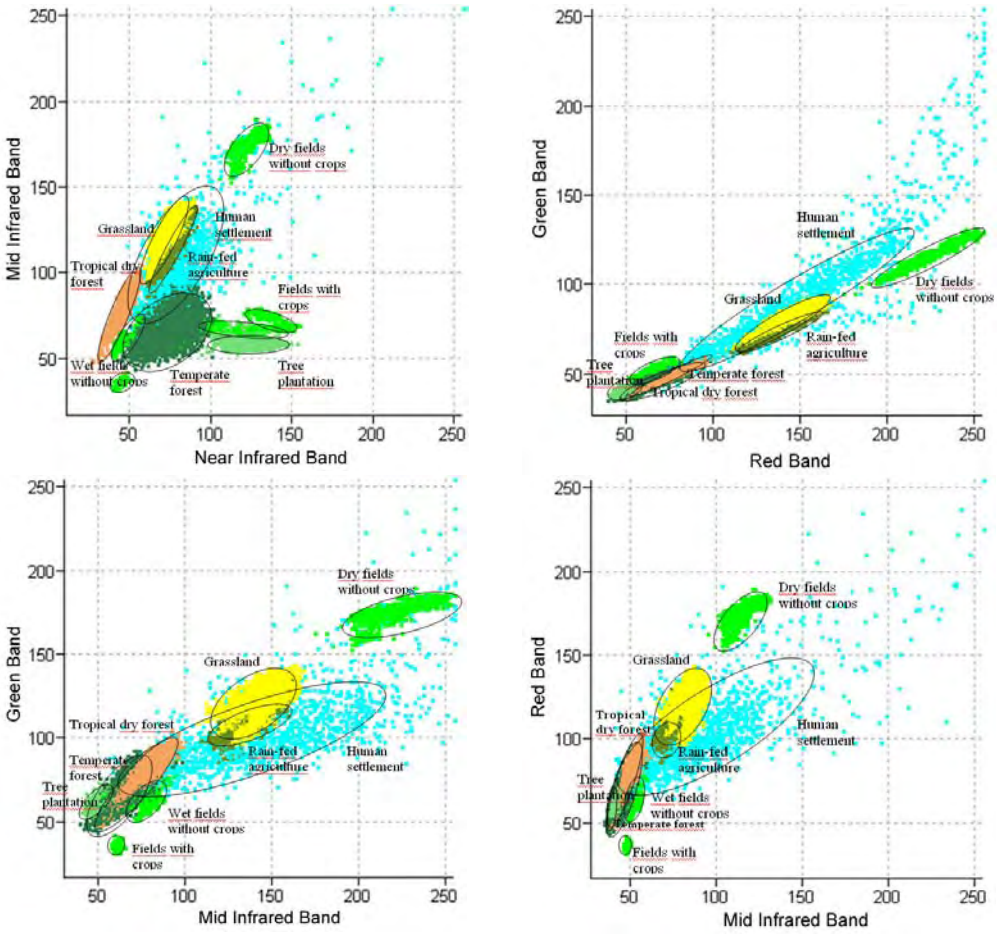


Figure 4-4 (continuation): *Feature space for spectral bands combinations of green, near infrared and red, near infrared, in the example of SPOT-5 image.*

Altogether there are seven land cover types present in the study area: 'irrigated agriculture', 'rain fed agriculture', 'grassland', 'tropical dry forest', 'human settlement', and 'orchards'. Figure 4-4 shows that there is overlapping in different extent between land cover classes in all the combinations of spectral feature space. There are three types of irrigated agriculture which appear distinctively in the image: agriculture fields with crops, dry agriculture fields without crops, and wet agriculture fields without crops. One type of irrigated agriculture-wet fields without crops-significantly overlap with tropical dry

forest due to that they share very similar spectral signature in most of the spectral bands. Also, class 'rain fed agriculture' and 'grassland' overlap evidently and both of them overlap with the class human settlement which is a spectrally very heterogeneous land cover class and in this study area some part of it is constructed from natural materials, causing substantial problems in spectral detection. Figure 4-4 showed that the near infrared band is important in distinguishing most of the land cover types, for example, in all the spectral feature combinations with near infrared band, the two classes irrigated agriculture and orchards are distinguishable, while they are not distinguishable when the near infrared band is not present.

4.4.2 Pixel-based classification results

The SPOT-5 image and the simulated images were classified with both pixel based MLC and pixel based NN classifiers, based on the selected training samples for the seven interested land cover classes. The classified images were firstly evaluated visually and it revealed that for the original SPOT-5 image with 10 meter spatial resolution, the classification has a strongly speckled result which gives it a "blurred" appearance. The class human settlement was mis-classified seriously although its training samples were greatly reduced to only one sample. Results show that with the increasing of the spatial resolution, the speckled-appearance problem is becoming less serious. Although the accuracy is decreasing, since in the very coarse spatial resolution image, the mixed pixels were the majority in the image and some classes, for example rain fed agriculture, orchards, and human settlement, having small areas in the image and thus are wrongly presented by the coarse spatial resolution image. Images generated by mean filtering method produced more homogeneous appearance in the classified images.

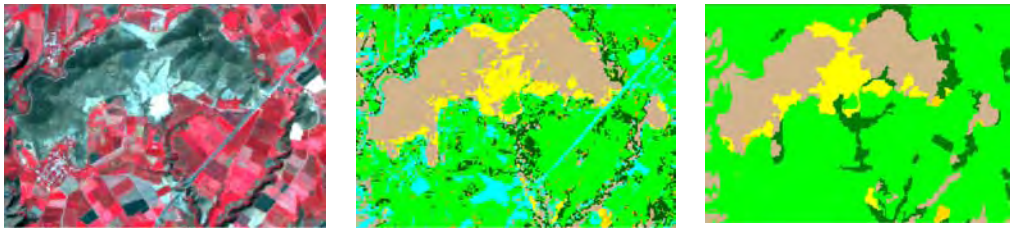
4.4.3 OBIA results

OBIA was carried out for the SPOT-5 image, two sets of simulated images, Landsat-7 ETM+, and MODIS image. Firstly, image segmentation was performed. The parameters to guide the segmentation process were explained in method section 3.4. For SPOT-5, the segmentation parameters were: scale

factor 20, colour factor 0.8, and smoothness 0.5 and it resulted in 23759 objects. For the two simulated images of 30m spatial resolution and Landsat-7 ETM+, the segmentation parameters settings were: scale factor 10, colour factor 0.8 and smoothness 0.5; it resulted in 10814 objects for the one by mean-filtering and 11721 objects for the one by cubic-filtering, and 7209 objects for Landsat-7 ETM+ image. For the two simulated images of 100m spatial resolution, the parameters for segmentation were: scale factor 5, colour 0.8 and smoothness 0.5; it resulted in 4696 objects for the one by mean-filtering and 5388 objects for the one by cubic-filtering. And for the two images with 250m spatial resolution and MODIS image, the parameter settings for image segmentation were: scale factor 3, colour 0.8 and smoothness 0.5; it resulted in 2265 and 3532 objects for the one by mean-filtering and the one by cubic-filtering, respectively, and 3075 object for MODIS image. Those parameter settings were decided based on visual checking that the produced segments represent optimally the primitive earth objects. The segmentation results showed that with the same parameter settings, simulated image by cubic-filtering produced more objects than those of images generated by mean-filtering, due to that the cubic-filtering retains the local contrast in the image while mean-filtering smoothes the contrast. The segmented images are classified by standard NN classifier using the same set of training samples as were used for pixel-based classification.

4.4.4 The comparison of pixel-based and OBIA results

Both pixel-based and OBIA results were evaluated with independent ground data which were generated by interpreting total 420 randomly distributed points. With the increasing of the spatial resolution (pixel size), ground data for 10m pixel may not be “true” for the property of the larger sized pixels. In order to evaluate the classifications of the images with different spatial resolution, the same number of ground data in the same location was interpreted for images with 10m, 30m, 100m, 250m spatial resolution, also for images of Landsat-7 ETM+ and MODIS. So altogether there are six sets of ground data to evaluate the accuracy of the classifications. The accuracies of the classification results were presented in figure 4-6 and 4-7.



simulated 30m | result by pixel-based | result by object-based

Figure 4-5: *Pixel-based and object-based classifications to a simulated image with 30m resolution. It showed that the object-based image analysis obtained results more homogeneous and closer to human vision.*

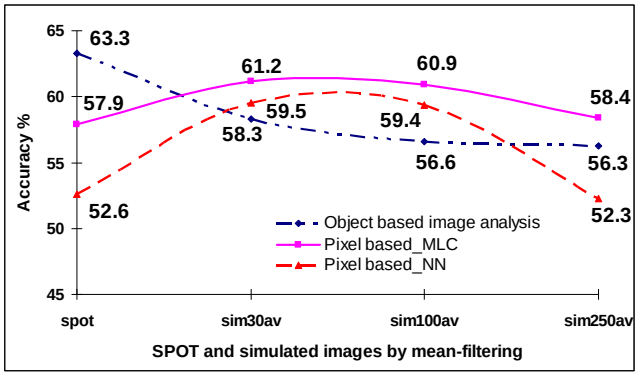


Figure 4-6: *Classification accuracies in function of image spatial resolution.*

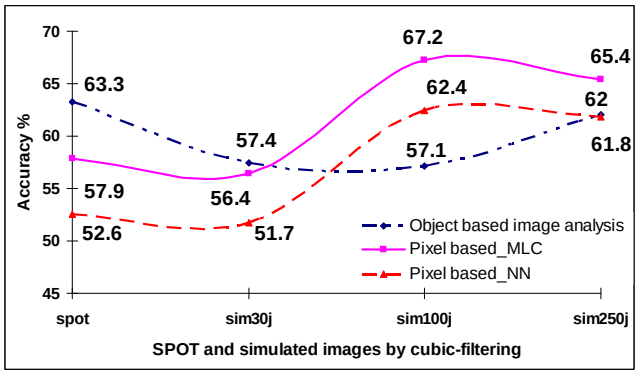


Figure 4-7: *Classification accuracies in function of image spatial resolution.*

In figure 4-6 and 4-7, the accuracy of the classifications using images produced

by cubic-filtering tends to be higher than using images produced by mean-filtering, except for simulated images with 30m spatial resolution. By averaging the pixel values of the windows, the mean-filtering produced images that have a smoother appearance than those by cubic-filtering. Images produced by cubic-filtering preserved the contrast in the original image and are more "heterogeneous" in appearance and in pixel values, and the produced images are closer to the images taken by the satellite sensors. If the accuracy is evaluated using homogeneous areas, images produced by mean-filtering have the tendency of optimizing the results. However, the classification accuracy here was evaluated by points and thus the classification results of images by mean-filtering did not present higher accuracy, instead, images by cubic-filtering which preserved the local contrast in the image obtained higher accuracy. As for the images with 30m spatial resolution, mean filtering clustered the pixels with similar spectral values, removed the spectral noises in the high spatial resolution images, while did not remove important local spectral contrast in the image, and thus obtained higher accuracy than images by cubic-filtering. Further observing figure 4-6 and 4-7, we can see that in the case of the two sets of simulated images, classification accuracies by pixel-based MLC are higher than those produced by pixel-based NN classifier, and with SPOT-5 and simulated cubic-filtering image, object-based image analysis obtained higher accuracy than those obtained by pixel-based MLC and pixel-based NN methods. This result showed that first, pixel-based MLC is indeed a well established pixel-based method which is able to obtained higher classification accuracy; second, for SPOT-5 image, object-based image analysis has advantage over pixel-based one regardless of pixel-based classification methods used. By classifying object which is a group of homogeneous pixels, object-based image analysis produces classification results closer to human interpretation results, free of speckled appearance, and with comparatively higher accuracies. Results on the simulated images by cubic-filtering showed that images with 100m spatial resolution obtained higher accuracy than that of the images with higher spatial resolution. This can be explained using the finding of Kan *et al*. (1975), that "theoretical justification and experimental verification support the finding that classification accuracies for low resolution data could be better than the accuracies for data with higher resolution. The increase in accuracy is due to the reduction of scene

in homogeneity at lower resolution". With the increasing of the spatial resolution of the images, OBIA obtained lower accuracy than that by pixel-based methods, which showed that the advantage of OBIA over the pixel-based one in classification accuracies is only represented by images with high spatial resolutions. Landsat-7 ETM+ and MODIS images were also classified and compared with classifications of images with the same spatial resolution. Pixel-based Landsat image classifications showed that by MLC it obtained an overall accuracy 50.1%, and NN obtained accuracy 48.8%. Object-based NN classification of Landsat image obtained an accuracy of 51.7% (figure 4-8).

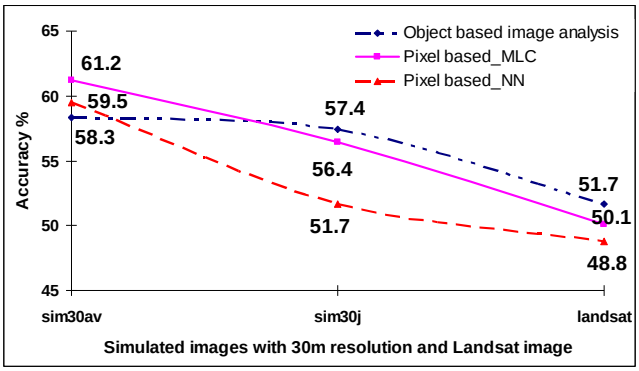


Figure 4-8: *Comparison of the accuracies of the classifications between original Landsat-7 ETM+ image and the simulated images with 30m spatial resolutions.*

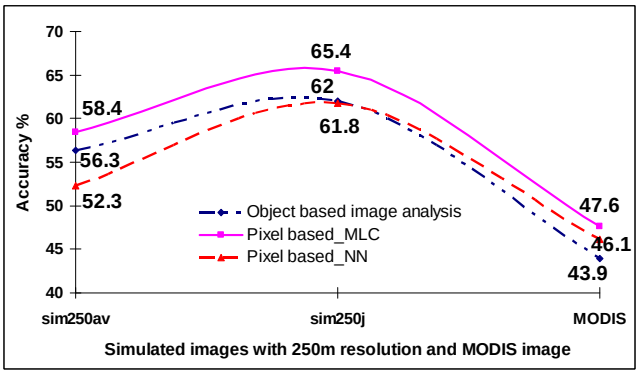


Figure 4-9: *Comparison of the accuracies of the classifications between original MODIS image and the simulated images with 250m spatial resolutions.*

For MODIS image, pixel-based MLC and NN classifiers obtained accuracy of

47.6% and 46.1% respectively, and object-based NN classification obtained an accuracy of 43.9% (figure 4-9). For Landsat image, object-based image analysis still obtained higher accuracy than those by pixel-based MLC and NN classifiers, while for MODIS, object-based method obtained accuracy lower than those by both pixel-based MLC and NN methods. This result proved again that for coarse spatial resolution images, object-based image analysis shows no advantage over the pixel-based one.

Figure 4-8 showed that Landsat image obtained results closer to the results of the images generated by cubic-filtering. This is because cubic-filtering method was developed by simulating the way of Landsat sensor taking the images. For the same reason, MODIS image did not present classification accuracy results closer to those by cubic-filtering method (figure 4-9).

Comparing with pixel-based image analysis, object-based image analysis has many advantages (Hay and Gastilla, 2006), such as: the way it classifies an image by partitioning it into objects is similar to the way humans comprehend the landscape; image-objects exhibit useful features (shape, texture, context relations with other objects) that single pixels lack; image-objects can be more readily integrated in vector GIS. This work performed pixel-based and object-based analysis to images with spatial resolutions from relatively fine to coarse. Both pixel-based and object-based image analysis results were evaluated with a set of stratified randomly sampled reference data comparing 420 points. Based on the result, a general conclusion was drawn that the advantage of object-based image analysis over the pixel-based one was only represented by images with higher spatial resolutions. Increased spectral variability within high resolution imagery confuses traditional pixel-based classifiers, while by object-based method, pixels with similar spectral information are firstly grouped into objects then those objects are analyzed. Images with medium to low spatial resolution have lower spectral variability and thus are easily handled by pixel-based method. As for the images with low spatial resolution, such as MODIS, by applying object-based image analysis, pixels belonging to different land cover types could possibly be grouped together thus are mis-classified and produce lower accuracy than that by pixel-based method. As stated before, the result drawn from this paper could never be absolute; there are many factors

which influence classification accuracy such as image data quality, the reliability of training data and testing/reference data, the accuracy assessment method adopted, among others. As to the accuracy assessment method, though accuracy evaluated by random points (a simple random or stratified random points) is generally recognized to be more trustworthy than that evaluated using homogenous areas of land cover types, it may not hold true when it comes to object-based image analysis results. Due to that single points/pixels are usually merged into surroundings by object-based method, and thus evaluated by single pixels, the accuracy of object-based image analysis could be under-evaluated, which may explain though the object-based image analysis results appear to be more appealing (figure 4-5), their accuracy results tend to be low. This is proved by checking the classification results of a simulated mean-filtering method with 30m spatial resolution (mean-filter30): 5.2% of the wrongly classified points for object based classification are coming from those isolated pixels. If we add this number to the already obtained object based classification accuracy, we get 63.5%, which is higher than those by both pixel based MLC and NN classifications of the image “mean-filter30” (figure 4-6).

4.5. Conclusion

In this paper, pixel-based and OBIA was performed on satellite images with different spatial resolutions. The coarser spatial resolution images are generated by degrading spatial resolution of a multi-spectral SPOT-5 imagery with a 10m spatial resolution. Two simulating methods were used: 1) mean-filtering which averages the pixel values in a certain sized window (3*3, 10*10, and 25*25) and, 2) cubic-filtering which is a modified Justice method that keeps the local contrast in the image during the interpolation and produced images closer to the “real” images. These two sets of images were classified by pixel-based MLC and NN classifier, and object-based NN classifier, respectively. Accuracy assessment results showed that for SPOT-5, and a simulated cubic-filtering 30m spatial resolution images, object-based image analysis obtained higher accuracies than those produced by pixel-based MLC and NN classifiers. With the increasing of the spatial resolution of the images, object-based image analysis did not show more advantage over the pixel-based ones. The classifications of the Landsat-7 ETM+ and MODIS images proved again the

object-based image analysis has advantage over the pixel-based one but the advantage represented by classification accuracy is only represented by high spatial resolution images.

Acknowledgement

The authors thank CONACYT-CONAFOR 2005-C02-14741 for supplying PhD fellowship to the first author during writing up of the paper. Many thanks go to Dr. Stephane Couturier for his valuable opinion in improving a previous version of this manuscript and his help in correcting the English grammar. Thanks also go to Msc Antonio Navarrete for elaborating figure 4-1 and figure 4-2.

References

Adams, R., and Bischof, L., 1994, Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16, pp. 641-647.

Baatz, M., and Schaepe, A., 1999. Object-oriented and multi-scale image analysis in semantic networks. 2nd International Symposium: Operationalization of Remote Sensing, 16-20 August, ITC, the Netherlands.

Baatz, M., Benz, U., Dehghani, S., Heynen, M., Holtje, A., Hofmann, P., Lingenfelder, I., Mimler, M., Sohlbach, M., Weber, M., Willhauck, G., 2004, eCognition User's Guide. http://www2.definiens.com/central/default.asp (the last accessed 07/March/2006).

Bastin, L., 1997, Comparison of fuzzy c-means classification, linear mixture modeling and MLC probabilities as tools for unmixing coarse pixels. International Journal of Remote Sensing, 18, 3629-3648.

Benz, Ursula C., Hofmann Peter, Willhauck Gregor, Lingenfelder Iris, Heynen Markus, 2004, Multi-resolution, object-based fuzzy analysis of remote sensing data for GIS-ready information. ISPRS Journal of Photogrammetry & Remote Sensing, 58, pp. 239-258.

Blaschke, T., Lang, S., Lorup, E., Strobl, J., and Zeil, P., 2000, Object based image processing in a GIS/remote sensing environment and perspectives

for environmental applications. In environmental information for planning, politics, and the public. Edited by A. Cremers and K. Greve. Metropolis-Verlag, Marburg, Germany, 2, pp. 555-570.

Blaschke, T., and Strobl, J., 2001, What's wrong with pixels? Some recent development interfacing remote sensing and GIS. GeoBIT/GIS, 6, pp. 12-17.

Borak, J.S., and Strahler, A.H., 1999, Feature selection and land cover classification of a MODIS-like data set for a semiarid environment. International Journal of Remote Sensing, 20, pp. 919-938.

Boschetti Luigi, Brivio Pietro A., Flasse Stephane, 2004, Pareto Boundary: a useful tool in the accuracy assessment of low spatial resolution thematic products. IEEE. http://www.gvm.jrc.it/tem/PDF_publis/2004/Boschetti%20etal_%20IGARSS04_Pareto_full%20text.pdf. (Last accessed 23/May/2005).

Campbell, J.B., 1996, Introduction to Remote Sensing (2nd ed.). London: Taylor and Francis.

Casals-Carrasco, P., Kubo, S., and Babu Madhavan, B., 2000, Application of spectral mixture analysis for terrain evaluation studies. International Journal of remote sensing, 21, pp. 3039–3055.

Canters, F., 1997, Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification. Photogrammetric Engineering and Remote Sensing, 63, 403-414.

Chintan A. Shah, Manoj K. Arora, Pramod K. Varshney, 2004, Unsupervised classification of hyperspectral data: an ICA mixture model based approach. International Journal of Remote Sensing, 25, pp. 481–487.

Dean, A.M., Smith, G.M., 2003, An evaluation of per-parcel land cover mapping using maximum likelihood class probabilities. International Journal of Remote Sensing, 24, pp. 2905–2920.

Definiens, 2006. Definiens professional User Guide 5. Definiens AG, Munich.

De Kok, R., Schneider, T., and Ammer, U., 1999, Object based classification and applications in the Alpine forest environment. In Fusion of sensor data, knowledge sources and algorithms, proceedings of the Joint ISPRS/EARSeL Workshop, 3-4 June 1999, Valladolid, Spain. International Archives of Photogrammetry and Remote Sensing, 32, Part 7-4-3 W6.

De Wit, A.J.W., and Clevers, J.G.P.W., 2004, Efficiency and accuracy of per-field classification for operational crop mapping. International Journal of Remote Sensing, 25, pp. 4091-4112.

Egawa, H., and Kusaka, T., 1988, Region extraction in SPOT data. Geocarto International, 3, pp. 25-32.

Evans Carolyn, Jones Ronald, Svalbe Imants, and Berman Mark, 2002, Segmenting multispectral Landsat TM images into field units. IEEE Transactions on Geoscience and Remote Sensing, 40, pp: 1054-1064.

Espindola, G. M., Camara, G., Reis, I. A., Bins, L. S., and Monteiro, A. M., 2006, Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *International Journal of Remote Sensing*, **27**, 3035-3040.

Flanders David, Hall-Bayer Mryka, and Pereverzoff Joan, 2003, Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. Canada Journal of Remote Sensing, 29, pp. 441-452.

Foody, G.M., 2002, Status of land cover classification accuracy assessment. Remote sensing of environment, 80, pp. 185-201.

Foody, Giles M., 2004, Thematic Map Comparison: Evaluating the Statistical Significance of differences in Classification Accuracy. Photogrammetric Engineering & Remote Sensing, 70, pp. 627-633.

Franklin, S.E., Peddle, D.R., Wilson, B.A., and Blodgett, C.F., 1991, Pixel sampling of remotely sensed digital imagery. Comput. Geosci, 17, pp. 759-775.

Fu, M., and Mui, C., 1981, A survey on image segmentation. Pattern Recognition, 13, pp. 3-16.

Gao, Y., Mas, J. F., Maathuis, B. H. P., Zhang, X. M., and Van Dijk, P. M., 2006. Comparison of pixel-based and object-oriented image classification approaches-a case study in a coal fire area, Wuda, Inner Mongolia, China. International Journal of Remote Sensing, 27, pp. 4039-4051.

Giada, S., T. De Groeve, Ehrlich, D., Soille, P, 2003, Information extraction from very high resolution satellite imagery over Lukole refugee camp, Tanzania. International Journal of Remote Sensing, 24, pp. 4251-4266.

Haralick, R. M., and Shapiro, L. G., 1985, Image segmentation techniques.

Computer Vision Graphics and Image Processing, 29, pp. 100-132.

Hofmann, P. 2001a. Detecting buildings and roads from IKONOS data using additional elevation information. In: GeoBIT/GIS 6, pp. 28-33.

Hofmann, P., 2001b, Detecting urban features from IKONOS data using an object-oriented approach. In: Remote Sensing & Photogrammetry Society (Editor): Proceedings of the First Annual Conference of the Remote Sensing & Photogrammetry Society 2001, pp. 28-33.

Integrated Land and Water Information System (ILWIS), 2005. International Institute for Geo-Information Science and Earth Observation (ITC). The Netherlands.

Janssen, L.L.F., and van der Wel, F.J.M., 1994, Accuracy assessment of satellite derived land-cover data: A review. Photogramm. Eng. Remote Sensing, 60, pp. 419-426.

Justice, C.O., Markham, B.L., Townshend, J.R.G., and Kennard, R.L., 1989, Spatial degradation of satellite data. International Journal of Remote Sensing, 10, 1539-1561.

Kan, E.P., Ball, D.L., Basu, J.P., and Smelser, R.L., 1975, Data resolution versus forestry classification and modeling. Symposium on Machine Processing of Remotely Sensed data, June 3 – 5, 1975, the Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana. IEEE Catalog No. 75CH1009-0-C.

Latty, R. S., 1984, Scene segmentation through region growing. In Proceedings, Tenth Symposium on Machine Processing of Remotely Sensed Data, West Lafayette, IN, 305-314.

Lillesand, M., Kiefer, W.R., 1994, Remote sensing and image interpretation, Third Edition, John Wiley & Sons, Inc., New York, 750 pp.

Lobo A., 1997, Image segmentation and discriminant analysis for the identification of land cover units in ecology. IEEE Transactions on Geoscience and Remote Sensing, 35, pp. 1136-1145.

Ouattara, T., Gwyn, Q.H.J., and J.-M. M. DUBOIS, 2004, Evaluation of the runoff potential in high relief semi-arid regions using remote sensing data: application to Bolivia. International Journal of Remote Sensing, 25, pp. 423–435.

Pal, Nikhil R. and Pal, Sankar K., 1993, A review on image segmentation techniques. Pattern Recognition, 26, pp. 1277-1294.

Pizzolato A.N., Haertel, V., 2003, on the application of Gabor filtering in supervised image classification. International Journal of Remote Sensing, 24, pp. 2167-2189.

Smits, P.C., Dellepiane, S.G., & Schowengerdt, R.A., 1999, Quality assessment of image classification algorithms for land-cover mapping: a review and proposal for a cost-based approach. International Journal of Remote Sensing, 20, 1461-1486.

Stehman, S.V., 1997, Selecting and interpreting measures of thematic classification accuracy. Remote sensing of Environment, 62, pp 77-89.

Stehman, Stephen V and Czaplewski Raymond L, 1998, Design and analysis for thematic map accuracy assessment: fundamental principles. Remote Sensing of Environment, 64, pp. 331-344.

Stuckens, J., Coppin, P. R., and Bauer, M. E., 2000, Integrating contextual information with per-pixel classification for improved land cover classification. Remote Sensing Environment, 71, pp. 282-298.

Tilton, J. C., 1996, Hybrid image segmentation for earth remote sensing data analysis. In Proceedings, International Geoscience and Remote Sensing Symposium, Lincoln, NE, 703-705.

Zhang, Y., and Maxwell, T., A Fuzzy Logic Approach to Supervised Segmentation for Object-Oriented Classification. In ASPRS 2006 Annual Conference, Reno, Nevada, May 1-5, 2006.

Zhou, Q. and Robson, M. 2001, Contextual information is ultimately necessary if one is to obtain accurate image classifications. International Journal of Remote Sensing, 22, pp. 3457–3470.

CHAPTER 5

A SEMI-AUTOMATIC FEATURE EXTRACTUIB ALGORITHM AND HIERARCHICAL NETWORK STRUCTURE FOR OBJECT-BASED IMAGE ANALYSIS[1]

Abstract. Object-based image analysis (OBIA) uses object's spectral and spatial information such as size, form, shape and texture information, and information pertaining to an object's sub-or super-objects if a multi-level image object hierarchy has been created. With all these available features, it is critical the proper selection of features and thresholds to differentiate the classes of interest. Separability and Threshold (SEaTH) algorithm is a semi-automatic feature extraction algorithm proposed by Nussbaum *et al*. (2006). It calculates the separability and thresholds of object-classes for any number of given features based on a statistical approach. With it, the time-consuming trial and error practice for seeking significant features and proper thresholds can be avoided. SEaTH algorithm can also help to minimize human involvement in classification steps and speeds up the classification process when huge data is to be dealt with, however, it could be at the expense of accuracy. This paper performed land cover classification using object based image analysis on

[1] Based on the content of this chapter, a paper titled "object oriented image analysis for a forest area land cover mapping" was presented as poster in 5th International Symposium Spatial Data Quality 2007-modelling qualities in space and time, 13-15, June, 2007, ITC, Enschede, the Netherlands.

Landsat-7 ETM+ imagery in a mountainous area with complex land cover classes. In the definition of the classifiers of the land cover classes, the used features and thresholds were calculated by SEaTH algorithm. Eight land cover classes were produced: temperate forest, tropical dry forest, orchards, grassland, irrigated agriculture, rain-fed agriculture, lava flow, and human settlements. In this chapter, it is differentiated between object primitives and object of interest. Object primitives are building blocks in object based image analysis and are produced by image segmentation; object of interest here corresponds to the earth objects.

Key words: Multi-resolution segmentation, Object-based image analysis, Separability and thresholds.

5.1. Introduction

Traditional pixel-based image processing algorithms produce highly freckled results in classifying land covers in heterogeneous areas and are limited in especially detecting small objects, particularly in analyzing high-resolution (HR) imagery such as Quick bird or IKONOS. The resulting thematic maps bear either very general land-cover information, or else detailed information with limited accuracies (Franklin *et al*. 2003). OBIA seems to be more precise and meaningful. The development of OBIA stems primarily from the desire to use the important semantic information necessary to interpret an image, which is not presented in single pixels but rather in meaningful objects and their mutual relations. In particular in OBIA, homogeneous image objects at a chosen resolution are first extracted and subsequently classified. In addition to spectral related information, OBIA uses information based on object size, shape, and context, and information pertaining to an object's sub-or super-objects if a multi-level image object hierarchy has been created (Shackelford and Davis 2003). This extra information considerably extends the possibilities for image analysis. OBIA intends to do what an image interpreter does: recognizing the colour, shapes, textures and coherent regions presented within an image and associating meaningful objects and their contextual relations.

A feature represents certain information concerning objects of interest. Feature

identification is a critical part of OBIA. It is important to identify the characteristic, significant features for objects of interest among the large number of possible features. SEparability and Threshold (SEaTH) algorithm calculates the separability and thresholds of object-classes for any number of given features based on a statistical approach and thus it avoided the time-consuming trial and error practice for seeking significant features and proper thresholds.

This paper investigates the performance of OBIA for mapping land-covers in a mountainous forested area. OBIA is performed in eCognition (Definiens, 2006) with multi-resolution segmentation, hierarchical network classification with membership function as classifier in which the rule-base was constructed with objects' features and thresholds calculated/extracted by SEaTH algorithm.

5.2. Methods

5.2.1. OBIA in eCognition

OBIA operates on objects created by image segmentation. Image segmentation is a kind of regionalization, which delineates objects based on a certain homogeneity criteria, and at the same time requires spatial contingency.

OBIA in eCognition (Definiens 2006) comprises two parts: multi-resolution segmentation and context-based classification. Multi-resolution segmentation allows generating image objects on an arbitrary number of scales taking into account criteria of homogeneity in colour and shape. Additionally, the created segments/objects are embedded into a hierarchical network in which each object knows its neighbouring objects in horizontal and vertical direction (figure 5-1) (Baatz and Schape 1999). Classification in eCognition can be based on a fuzzy nearest neighbour classifier or membership function. The nearest-neighbour (NN) classifier has long been used in pattern recognition, exploratory data analysis, and data mining problems (Tahir *et al*. 2006). A vital consideration in obtaining good results with this classifier is the choice of distance function, and correspondingly which features to consider when computing distances between samples.

5.2.1.1. Multi-resolution image segmentation

Image scale is an important concept for meaningful image analysis. Therefore, eCognition provides segmentation on multiple scales. Image segmentation in eCognition is a multi-resolution, region-merging technique starting with one-pixel objects. The segmentation procedure can be applied on the pixel level or an image object level. Image objects are extracted from the image in a number of hierarchical segmentation levels, and each subsequent level yields image objects of a larger average size by combining objects from a level below, which represents image information on different scales simultaneously (figure 5-1).

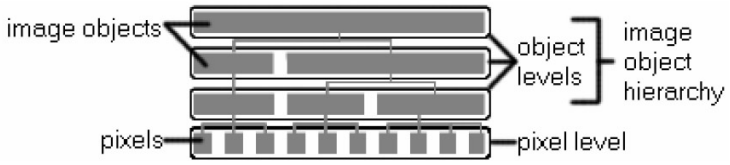


Figure 5-1: *Image object hierarchy. Source: Definiens professional 5 user guide.*

There are several parameters to guide the segmentation result. Scale parameter is an abstract term which determines the maximum allowed heterogeneity for the resulting image objects. With a certain scale parameter, objects are formed based on spectral similarity and shape characteristics of the resulting object. Spectral similarity is defined by "colour" criterion and "shape" characteristics are represented by both smoothness and compactness criteria.

Colour criterion defines the weight the spectral values of the image layers contribute to image segmentation, as opposed to the weight of the shape criterion. There is the relation between colour and shape criteria: colour + shape = 1. Maximum colour criterion 1.0 results in objects spectrally most homogeneous; while with a value less than 0.1, the created objects would not be related to the spectral information at all. Smoothness is to optimize image objects with regard to smooth borders, and compactness is to optimize image objects with regard to compactness (Baatz *et al*. 2004). The resulting objects

also depends on the image data, for a given set of segmentation parameters, heterogeneous image data result in smaller image objects than homogeneous image data.

5.2.1.2. The fuzzy classifier

Image objects can be described and classified by an extensive variety of features that include spectral related information such as mean, standard deviation, etc. and texture, form, context properties in several forms. The classification can be based on a fuzzy nearest neighbour classifier (NN) or fuzzy membership functions, or a combination of both. The variety of object features can be used either to describe fuzzy membership functions, or to determine the feature space for NN (Baatz and Schape 1999).

Object based NN classifier is based on fuzzy logic. Fuzzy classification is a "soft" classifier which takes into account uncertainties in sensor measurements, parameter variations due to limited sensor calibration, vague (linguistic) class descriptions, and class mixtures due to limited resolution.

Fuzzy classification consists of n-dimensional membership degrees, which describes the degree of class assignment of the considered object to the n considered classes.

$$f_{class,obj} = \left\lfloor \mu_{class\_1}(obj), \mu_{class\_2}(obj), ... \mu_{class\_n}(obj) \right\rfloor \qquad (5.1)$$

Crisp classification gives only the highest membership degree information, whereas the fuzzy classification result contains all information about the overall reliability, stability and class mixture. Fuzzy logic replaces the two Boolean logical statements "true" and "false" by the continuous range of (0…1) where 0 means false and 1 means true and all values between 0 and 1 represent a transition between true and false. Avoiding arbitrary shape thresholds, fuzzy logic is able to approximate real world in its complexity much better than the simplified Boolean systems do. Fuzzy logic can model imprecise human thinking and can represent linguistic rules. Hence, fuzzy classification systems

are well suited to handle most sources of vagueness in remote sensing information extraction. The mentioned parameter and model uncertainties are considered by fuzzy sets, which are defined by membership functions. Fuzzy classification requires a complete fuzzy system which consists of three main steps 1) fuzzification of input variables resulting in fuzzy sets, 2) fuzzy logic combinations of these fuzzy sets and 3) de-fuzzification of the fuzzy classification result to get the common crisp classification for land-cover land-use map production.

5.2.1.3. Hierarchical network classifier

Multi-resolution segmentation procedures provided by eCognition operate on arbitrary levels in a strong hierarchical network. To guarantee a definite hierarchy over the spatial shape of all objects the segmentation procedures follow two rules: (1) Object borders must follow borders of objects on the next lower level. (2) Segmentation is constrained by the border of the object on the next upper level.

Objects created in different scales can be linked together in a hierarchical network like the one displayed in figure 5-1. The hierarchy network has several advantages for image analysis. Since structures of different scales are represented simultaneously and thus can be classified in relation to each other. Different hierarchical levels can be segmented based on different data, an upper level, for example, can be built based on thematic information, whereas a lower layer is segmented using remote sensing data. Classifying the upper level, each object can be analyzed based on the composition of its classified sub-objects. By means of this technique different data types can be analyzed in relation to each other.

5.2.1.4 Important object features

Since regions in the image provide much more information than single pixels, there is a large number of different image object features for measuring colour, shape, and texture of the associated regions. Even more information may be extracted by taking the network structure and the classification of the image

objects into account. Important examples of this type of features are the “rel. border to neighbouring objects” of a given class and “number of subobjects” of a given class. Equation 5.2 gives an example of the ratio of the mean values of two input channels A and B.

$$f_{rAB} = \frac{\frac{1}{n}\sum_{n} P_A(x_n)}{\frac{1}{n}\sum_{n} P_B(x_n)} \qquad (5.2)$$

In which n is the number of pixels; x is the object, p(x) is the value of pixel at location x.

Besides spectral related features, there is also shape feature. The closer object primitives are to objects of interest, the more image-object shape features such as size, length or number of edges can be used. Advanced shape features can be derived from object polygons and object skeletons. Polygons are vector objects which provide new and more detailed information for the characterization of image objects by their shape. They are also needed for the visualization and export of image object outlines. Skeletons, which describe the inner structure of an image object, help to describe an object’s shape more accurately. Since these features are independent of sensor characteristics they are robust versus sensor calibration and illumination conditions.

As for the texture feature, the Grey Level Co-occurrence Matrix (GLCM) is used. GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image. It is used for a series of “second order” texture calculations. The first order texture measures are statistics calculated from the original image values, like variance, and do not consider pixel neighbour relationships. The second order measures consider the relationship between groups of two (usually neighbouring) pixels in the original image. Third and higher order textures (considering the relationships among three or more pixels) are theoretically possible but not commonly implemented due to calculation time and interpretation difficulty.

As for the topological object features, due mainly to the object hierarchical network, context features are provided. Within one scale, relations to neighbouring objects can be evaluated, whereby the range of the neighbourhood can be defined as parameter. Between image scales, hierarchical relations can be explored, where the distance of scales can be adjusted using distance scale parameter. The object hierarchical network provides additional object features and the characterization of an image object can be based on its sub-objects using texture and line analysis; class-related features can also be used to define the object for example relationships to classified sub-objects, such as the relative area of image objects assigned to a certain class, e.g. if an urban area on higher level contains many sub-objects classified as houses, this urban area can be described as dense vs. less dense areas. The characterization of an image object can also be based on its super-objects, e.g. houses belonging to a super object urban can be classified as urban houses, whereas houses belonging to a super objects rural area can be classified as cottages or special buildings.

Another very important type of features is semantic features. These higher order features are available after a first classification of image objects. They allow describing a park as forested region within urban area or shore regions as adjacent land regions to water. These semantic features reduce ambiguities, allow land-use classification in addition to pure land-cover classification and thus lead to a first step of scene understanding.

5.2.2. SEaTH algorithm

Generally, a semantic class can be described by its characteristic features and their distribution in the feature space. Using OBIA to analyze an image, there are many features available to describe the objects of interest. Therefore it is necessary to determine the prominent features for each object class for the succeeding image analysis (Nussbaum *et al*. 2005). Marpu *et al*. (2006) and Nussbaum *et al*. (2006) proposed a feature analyzing tool SEaTH (SEparability and THresholds) which identifies the characteristic features with a statistical approach based on training objects. This tool measures the statistical distance (Jeffries-Matusita distance), which involves the covariance matrix in the

comparing of the separability, of the features of two interested classes. The measure result is in interval of [0, 2], and J = 0 implies that the two distributions are completely correlated and J = 2 implies that the distributions are completely uncorrelated. The features which have very high J value are the optimum features which characterize the classes (Nussbaum *et al*. 2006, Marpu *et al*. 2006). By this method, first of all, samples of the interested classes are selected in eCognition and the statistics of those samples are exported. The separability of the classes represented by features J distance is then calculated. Based on the calculated results, the features with the high J distance are favorable feature to differentiate between land-cover classes. The selected favorable features can then be used to build the rule base for image classification in eCognition. For the details of SEaTH algorithm see Nussbaum *et al*. (2006).

By SEaTH algorithm, the aim is to classify the image into the interested land-cover classes. Because it is a parametric approach, the assumption is that all the random variables encountered during the processing can be approximated as Gaussian distributions. The image analysis procedure can be summarized in the following 5 steps (Marpu *et al*. 2006, Nussbum *et al*. 2006):

1) Image preparation/pre-processing. It is an optional step, which is to ensure proper segmentation (partition the image into contiguous and continuous groups of pixels based on a certain criteria) by means of creating new image layers or enhancing existing image layers. Image segmentation delineates image objects according to a certain criteria of homogeneity, and at the same time, requires spatial contingency (Lang and Blaschke 2006). Segmentation provides building blocks of OBIA. Different image scenario requires different pre-processing steps. In some cases, using averaging filter on the image can help to extract image objects. In analyzing images with land cover types such as vegetation, it is beneficial to calculate the Normalized Difference Vegetation Index (NDVI) and use it in image segmentation step.

2) Sample collection after segmenting the image into primitive objects. This can be done either manually by selecting samples in the image based on human interpretation or statistically by selecting specific regions in the concerned

image histograms. For example if the objects of class of interest are characterized by bright regions compared to other objects in the image, then 2-5% of the objects in the image histogram which have high mean values can be selected as samples for the class of interest, and 2-5% of image objects which have low mean values are assigned as samples to the other class.

3) Features are the characteristics defined for an object. For example mean value, standard-deviation, length, area, and texture, etc. are all the possible features of an object (Baatz *et al*. 2004). This step is to identify the optimal object features in order to separate land cover classes properly. For some features, the probability distribution of the object classes has small or large overlap, and the object classes have partial separability using these features. The features in which the object classes have large overlap are not optimal for differentiating them from other classes and it could cause serious misclassifications when these features are used in classifying the object classes (figure 5-2). Features in which object classes have no overlap are well suited for separating these object-classes by supplying complete separability and those features are optimal features. The optimal features can be identified visually by observing displayed feature values. As image objects can be characterized by a large number of features, it is often a laborious task to identify them visually. To intuit what features can be possible candidates needs a good understanding of the analyzed images and the classes of interest.

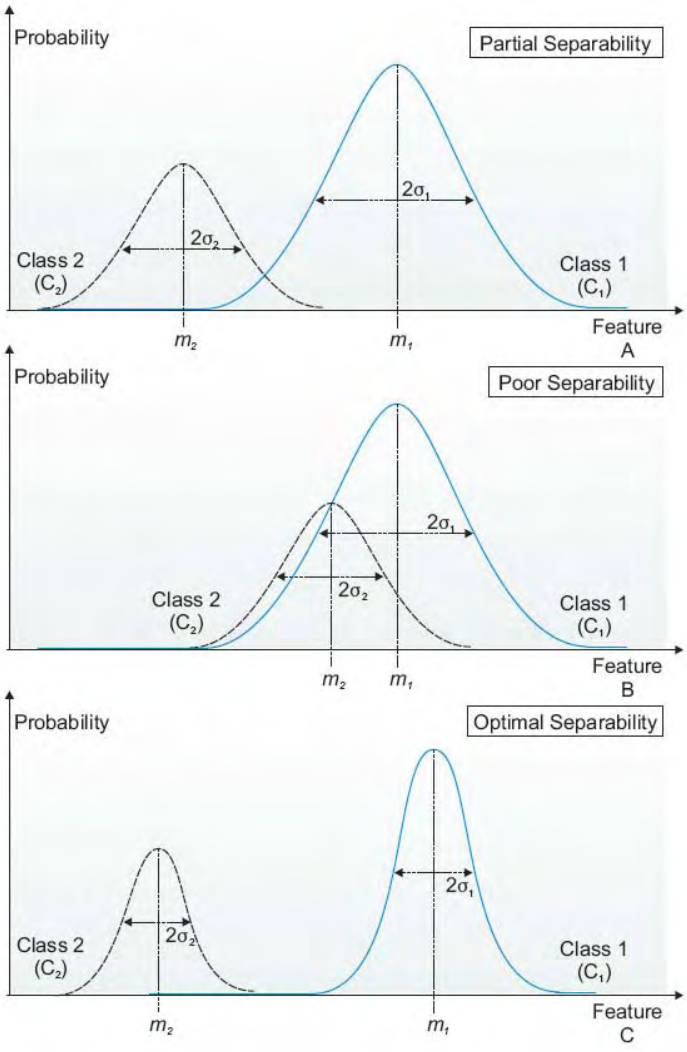


Figure 5-2: *Probability distributions. Source: Nussbaum et al. 2006.*

Characteristic features identification is through probability density estimation. Based on representative samples for object classes, the probability distribution for each class can be estimated and used to calculate the separability between two object classes. Under the assumption of normal probability distributions, the Bhattacharyya distance B can be used as a suitable separability measure. B is justified as a measure of separabillity from the Bayesean decision rule for misclassification probability. For the derivation of the Bhattacharyya distance

sees (Bhattacharyya, 1943), or (Fukunaga 1990). For two classes (c1, c2) and one feature Bhattacharya distance is expressed as

$$B = \frac{1}{8}(m_1 - m_2)^2 * \frac{2}{\sigma_1^{\,2} + \sigma_2^{\,2}} + \frac{1}{2}\ln\left(\frac{\sigma_1^{\,2} + \sigma_2^{\,2}}{2\sigma_1\sigma_2}\right) \qquad (5.3)$$

where m_i and σ_i , i=1, 2, are the mean value and standard deviation, respectively, for the feature distributions of the two classes. If the means coincide, the first term in (5.3) vanishes, whereas the second term vanishes if the feature distributions of the two classes have equal variances.

And based on this, Jeffries-Matusita distance (Nussbaum *et al*. 2005) is calculated as

$$J = 2(1 - e^{-B}) \qquad (5.4)$$

Using Jeffries-Matusita (J), the optimal features can be identified automatically based on the statistical values of the selected object samples. Note: The reason to calculate the J distance is that its values are in the range of (0-2). Other distance measures may also be used in this step to identify the optimum features.

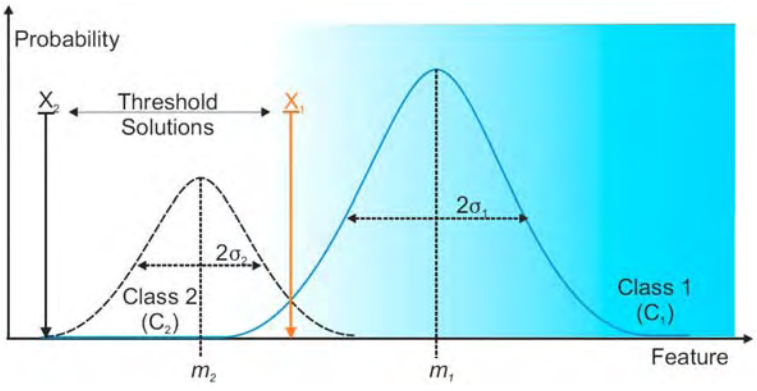


Figure 5-3: *Threshold identification. Source: Nussbaum et al. 2006.*

4) With respect to the means of the samples of two classes and in the feature space defined by the optimal features identified in the previous step, clusters are formed using the minimum distance criterion. This clustering is an approximation of the desired classification. To represent all the features on a common scale, a transformation has to be made on the feature values before clustering. This step transforms all the feature values to the range of (0, 1). For every object feature value F of a particular feature,

$$F_1 = F - F_{\min} \qquad (5.5)$$

$$F' = \frac{F_1}{F_{1\max}} \qquad (5.6)$$

$F_{\min}$ is the minimum of the object feature values of that feature, $F_{1\max}$ is the maximum of value F_1 obtained in the first step. F' is the transformed feature value of F.

5) Besides determining the optimal features separating the object-classes, it is essential to know the thresholds for the maximum separability to form the rule based for the classification. The optimal threshold is also calculated by SEaTH. A Gaussian probability mixture model of the form

$$p(x) = p(x/c_1)p(c_1) + p(x/c_2)p(c_2) \qquad (5.7)$$

is fit to the frequency distribution of a feature for two object classes c1 and c2, where p(x/c1) is a normal distribution with mean mc1 and variance $\sigma^2 c1$ and similarity for p(x/c2). The decision threshold which minimizes the error probability is obtained by solving

$$p(x/c_1)p(c_1) = p(x/c_2)p(c_2) \qquad (5.8)$$

for x. Taking logarithms,

$$\frac{1}{2\sigma^2 c_2}(x - mc_2)^2 - \frac{1}{2\sigma^2 c_1}(x - mc_1)^2 = \log\left[\frac{\sigma c_1}{\sigma c_2} * \frac{p(c_2)}{p(c_1)}\right] =: A \qquad (5.9)$$

Where $A = \log\left(\frac{\sigma_1}{\sigma_2} * \frac{n_2}{n_1}\right)$ (5.10)

the final thresholds for features to separate the two classes is then found based on Bayes' conditional probability principle. For classes C1 and C2, the threshold can be found as

$$x_{1(2)} / T = \frac{m_2\sigma_1^{\,2} - m_1\sigma_2^{\,2} + -\sigma_1\sqrt{m_1 - m_2^{\,2} + 2A(\sigma_1^{\,2} - \sigma_2^{\,2})}}{(\sigma_1^{\,2} - \sigma_2^{\,2})} \qquad (5.11)$$

Using formula (5.10), the thresholds of features can be calculated for the classes of interest, based on the distribution of objects classified in the previous step. Since the distributions are only partially separated, there will be some misclassifications when using this feature for classification of unknown object classes. Given the validity of the normal approximation assumption, SEaTH minimizes the misclassification number.

5.2.3. Image preparation by Principal component analysis (PCA)

PCA is a technique to reduce the spectral redundancy in multispectral data. It is often performed as a preprocessing procedure prior to automatic image classification. The purpose of this procedure is to compress the information contained in an original n-band data set into components or "new bands" fewer than n. The PCA data values are simply linear combinations of the original data values. The data along the first principal component have a greater variance or dynamic range than the data plotted against either of the original axes. For channels of multispectral data, the first principal component (PC1) includes the largest percentage of the total scene variance and succeeding components (*PC2, PC3, …PCn*) contain a decreasing percentage of the scene variance.

Furthermore, because successive components are chosen to be orthogonal to all previous ones, they contain uncorrelated data (Lillesand and Kiefer 2001). Used in the classification procedure, PCA data are normally treated simply as if they were original data. However, the number of components used is normally reduced to the intrinsic dimensionality of the data, thereby making the image classification process much more efficient by reducing the amount of computation required. For example, Landsat-7 ETM+ or TM data can often be reduced to just three principal components for classification purposes. PCA is particularly appropriate where little prior information concerning a scene is available.

5.3. The study area and data

The study area is located in Michoacán state, central west of Mexico, covering an area of approximately 58*60 km^2, within the longitude of 102° 00' W and 102° 32' W, and latitude of 19° 02' N and 19° 36' N (figure 5-4).

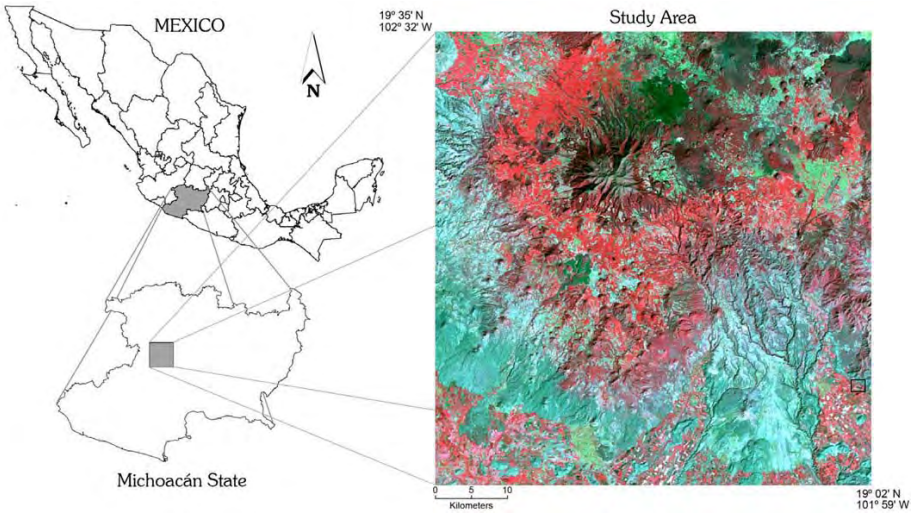


Figure 5-4: *The study area. Left side of the figure are two sketch maps indicating Mexico and Michoacán state where the study area is located; right side is the false colour composite of Landsat-7 ETM+ image with red, green, and blue bands of 4, 5, and 7.*

The available data are comprised of a Landsat-7 ETM+ image obtained on 16 February 2003, with 6 bands and a spatial resolution of 30m, a mosaic of 25 ortho-corrected photos taken in 1995 with 2 meters spatial resolution, and a land-cover map from the Forest Inventory generated in the year 2000. Eight land-cover types need to be classified: 'orchards', 'temperate forest', 'irrigated agriculture', 'rain-fed agriculture', 'lava flow', 'tropical dry forest', 'grass land', and 'human settlement'.

5.4. Results and discussion

PCA was carried out to the six Landsat-7 ETM+ spectral bands and three PCA bands were generated. NDVI was calculated based on bands 3 (red), and 4 (near infrared) and its value was calibrated into the range of 0-300. The three PCA bands, the calibrated NDVI, and the six Landsat spectral bands were used in the OBIA.

Two segmentation levels were created by multi-resolution segmentation using different parameter settings. In the first level image was segmented using both image layers and one thematic layer which contains 'human settlement' information. Including the information of a thematic layer in the segmentation influences the generation of image objects. In contrast to image layers, thematic layers contain discrete information, which means that related layer values can carry different information that is defined in the attached attribute list. To be able to clearly define the affiliation of an object to a thematic class given by the thematic layer, it is not possible to create image objects which belong to different thematic classes. To ensure this, the borders separating different thematic classes are restrictive for further segmentation whenever a thematic layer is used. For this reason, thematic layers can not be given different weights but can merely be selected for use or not. In this case, of course, the thematic layer which contains the settlements information was given the weight one to include it in the segmentation. Meanwhile, the image layers can also be given to different weights. If switch the weights for all image layers to 0, the produced image objects are based exclusively on thematic layer information and the obtained objects do not represent the averaging of layer

values. The purpose of the segmentation in this level was to create only settlement and non-settlement objects. In order to group all the objects which are not the settlement objects into one object for facilitating the object allocation, an extremely large scale factor 10000 was used and the segmentation created 'human settlement' objects and one non-settlement object (figure 5-5). This level acted as super level in class hierarchy.

Figure 5-5: *Segmentation using the thematic layer (level 1)*

A second segmentation level was created with the three PCA bands and the calibrated NDVI with parameters of scale factor 25, colour factor 0.7, compactness 0.5. The selection of appropriate segmentation factors was based on trial and error with the segmentation procedure until a satisfactory pattern was found. In this case, the produced segments are smaller than the objects of interest. The importance of visual inspection in segmentation result evaluation was addressed in Pal and Pal (1993) and Benz (2004) and it is still a common practice in segmentation parameters selection. The settings of all segmentation parameters are very much dependent on the image data, study area, and the desired land cover types, which gives the segmentation black box characteristics, though several groups are working on approaches to overcome this limitation (e.g. Espindola *et al*. 2006).

These two levels were differentiated using hierarchical membership function. The second level is the actually classification level (figure 5-6). The smaller objects (in level 2) are nested within larger ones (in level 1) and inherit any characteristics of a larger image object.

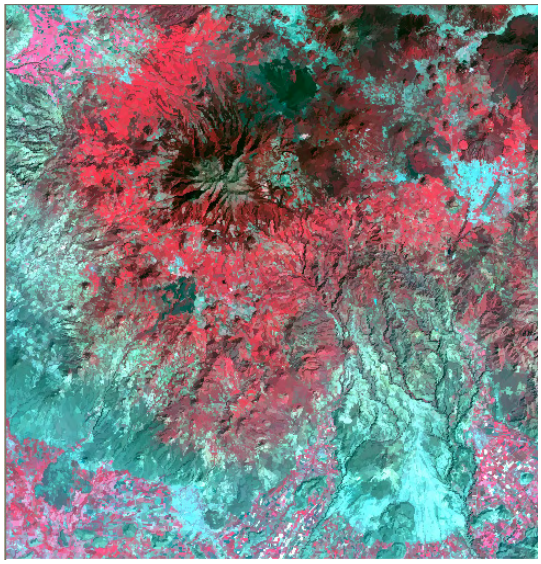


Figure 5-6: *Segmentation by scale factor 20, colour 0.7, and compactness 0.5 (level 2).*

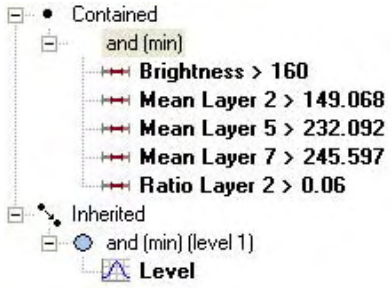
First the super-level was classified. Two child classes were created in this level: 'human settlement' and 'non human settlement'. These two classes can be classified either manually since there are only a few objects to be defined or by NN through defined samples for class human settlement and non-settlement. In this chapter, they were classified using NN classifier.

Eight classes were created in the classification level: 'human settlement', 'irrigated agriculture', 'rain-fed agriculture', 'grassland', 'tropical dry forest', 'orchards', 'temperate forest', and 'lava flow'. 'Human settlement' at this level was classified using the classification result from its super level "class related features, relation to super class 'human settlement'. The rest seven class use also class related features from the super level class 'non-settlement': "class

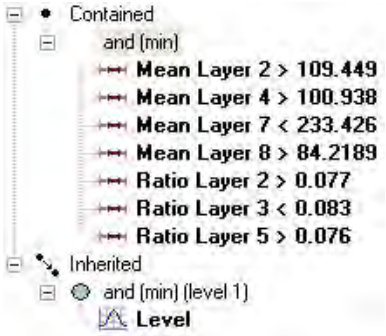
related features, relation to super class 'non-settlement'. Besides this class related feature, other features used were calculated using the SEaTH algorithm. For SEaTH analysis about 2-5% sample objects of each land cover class were selected as training data. Then 104 characteristic features including spectral related, shape, and texture features were selected for the training objects. Based on the training objects and selected features, SEaTH tool calculated, for every object class combination, the separability and thresholds for each of the 104 features. Based on the separability value, the "best" among 104 features are selected for each object class combination, and the features with maximum separability are used for classification. The selected features and thresholds based on SEaTH results for the land-cover classes were shown in the figure 5-7.

The calculated features and thresholds are presented in the "class description" in eCognition. A class description is composed of expressions which describe a class through fuzzy or crisp terms and are connected by logical operators. Logical operators are needed because usually there are multidimensional dependencies in the feature space and a logical combination of features is used to represent this condition. The most common operators are "and (min)" and "or (max)" which are used in the following way:

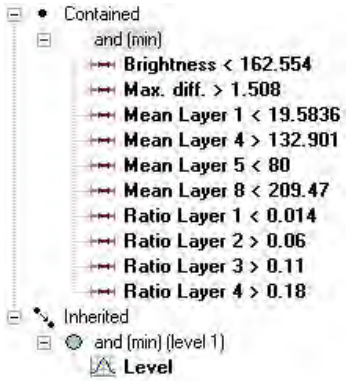
"And (min)": the membership of the output equals the minimum fulfillment of the single statements, which means that out of a number of conditions combined by the minimum operator, the lowest membership value is returned. "Or (max)": the membership of the output equals the maximum fulfillment of the single statements, which means the highest membership value is returned. For the minimum and maximum operators, only one statement determines the output, while for all other operators the values of all contained conditions contribute to the output. When creating a new class, the minimum operator "and (min)" is used by default. In the following, icon "contained" shows the classifier description that was inserted directly into the specified land cover class, and the icon "inherited" show the inherited classifier description from a higher classification level. Below the "contained" icon, logical expression "and (min)" is used, which returns the minimum of the fuzzy values.



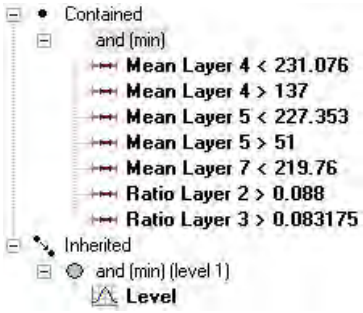
Dry fields without crops



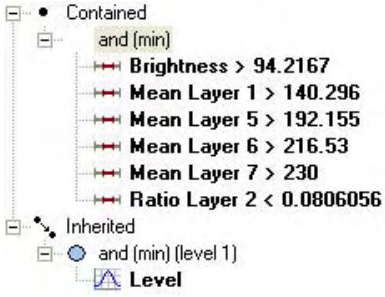
Fields with crops



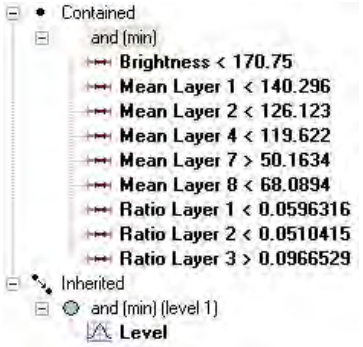
Dense temperate forest



Sparse temperate forest

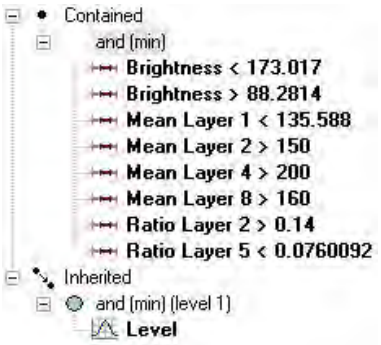


Grassland

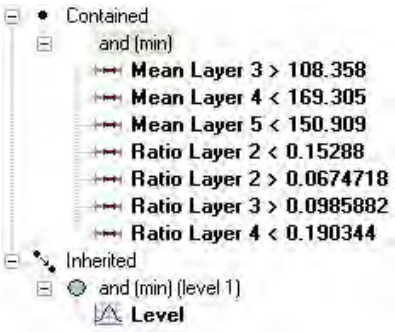


Lava flow

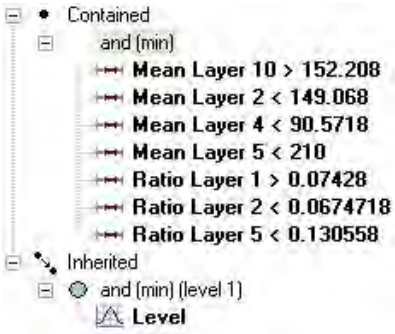
Figure 5-7: *The calculated features and thresholds for the land cover classification by SEaTH algorithm.*



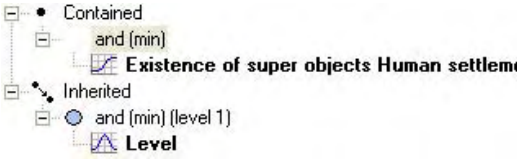
Orchards



Oyamel pine

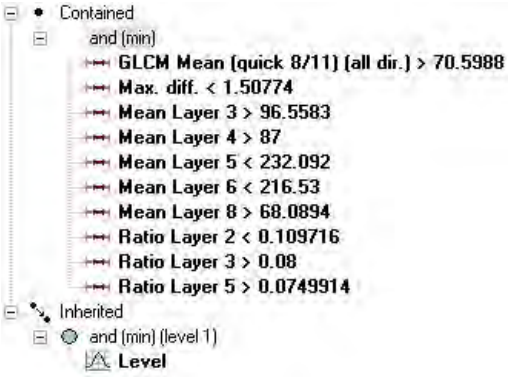


Rain fed agriculture



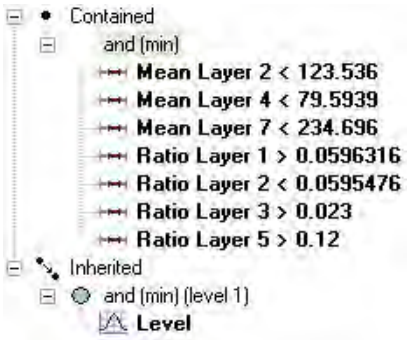
Settlement

Shadow

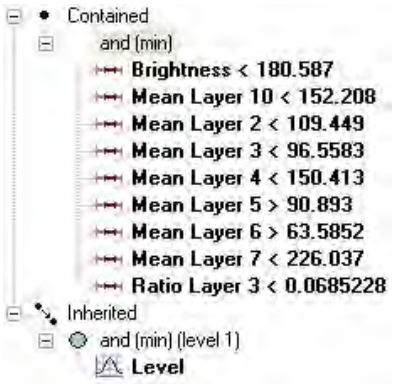


Sparse vegetation

Figure 5-7 (continuation): *The calculated features and thresholds for the land cover classification by SEaTH algorithm.*



Tropical dry forest



Wet fields without crops

Figure 5-7 (continuation): *The calculated features and thresholds for the land cover classification by SEaTH algorithm.*

In the classification scheme, class 'irrigated agriculture' has three sub-classes and 'temperate forest' has five sub-classes. Using SEaTH algorithm, the training objects were selected for each of the sub-classes and SEaTH calculated the features and threshold for the sub-classes. In eCognition, after the classification of the sub-classes, they can be grouped together as the classes of interest.

The classifier of the land cover classes were then defined with those features and thresholds and were classified. Most of the object primitives were classified into the designed land cover classes. The classification using the features and thresholds calculated by this SEaTH algorithm does not produce fuzzy results, and each classified object primitive has the full membership to its classified class. Due to the spectral variety of objects of interest, it is possible that the selected training samples do not always cover the spectral range and represent the characteristics of those objects, and thus there are also objects that were not classified in the classification result. Generally speaking, if there are a large amount of objects unclassified, the process of selecting the training objects and calculating the features and thresholds should be re-done; if there are relatively few objects unclassified, the calculated thresholds can be adjusted in order that the feature values of those objects are being included and thus they can be classified. First it needs to be clear which classes those unclassified objects should belong to, and then the object feature values are compared with

those desired classes. Although adjusting the calculated threshold could cause the mis-classification of other objects and it should be done carefully. When there are very few objects remaining unclassified, we can use the manual classification to allocate them to the correct classes.

In this case, there were only a reduced number of objects unclassified. By observing the characteristics of those objects and comparing them with the features and thresholds of the desired classes, the thresholds of the classes were adjusted and the objects were re-classified. The classification result is shown in Figure 5-8, and the legend of the produced classification is shown in figure 5-9.

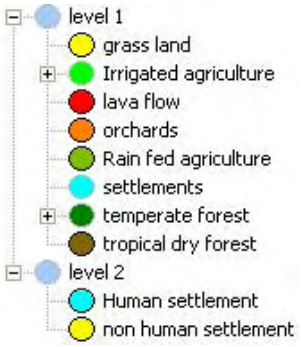


Figure 5-9: *Legend of SEaTH classification result (level 2)*

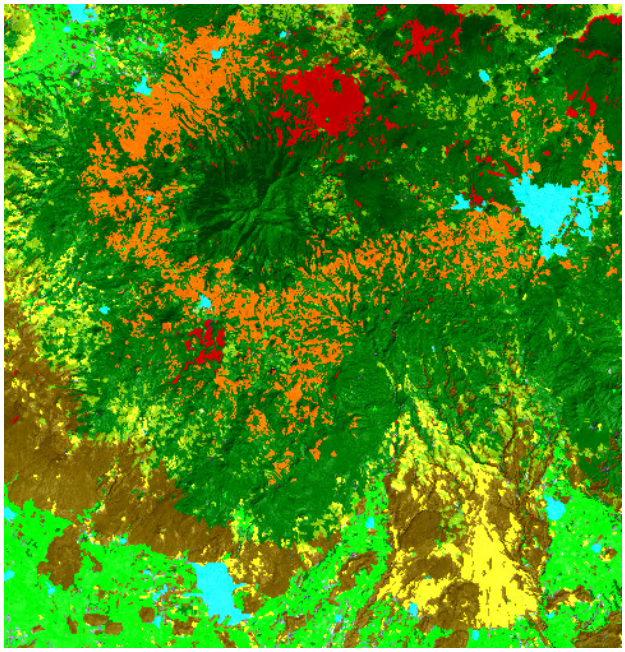


Figure 5-8: *Classification result of level 2.*

Accuracy assessment was carried out based on 600 stratified random points. 75 random points were put to each of the eight classes of interest in a land use map (2000). These points were interpreted based on the ortho-corrected photographs, land use map 2000, data from field survey, and Landsat-7 ETM+ images. To evaluate the accuracy of the object based analysis, error matrix was formed

based on the ground data and classification data and it is shown in table 5-1 (see attached file). Accuracy indices such as producer's and user's accuracies were calculated and it obtained an overall accuracy 78.7%.

5.5. Conclusions

Image objects can be characterized by a variety of spectral, spatial, texture, and contextual features. A comprehensive feature analysis is essential to work with image objects and the process of looking for the proper features could hinder the proper utilization of the strength of object based image analysis. Separability and Threshold (SEaTH) algorithm is able to evaluate statistically any number of given features for any number of object classes of interest. It utilizes the advantage of a wide feature basis of objects and provides the fundament for a successful classification. With the SEaTH algorithm, the optimum features which characterize a class can be identified to separate the objects of interest. The working process of SEaTH algorithm can be generally described as the following: first, training objects for the classes of interest were selected. If a class has several sub-classes, training objects should be selected for each of the sub-class. Then the feature values of those training objects are exported and the desirable features and thresholds are calculated based on the separability. For each class, SEaTH compares it to the other classes using each of the selected features. The features and thresholds were listed in the sequence of the separability values. Based on the calculation, features and thresholds with the highest separability-the highest Jeffries-Matusita (J) value, which is in the range of (0, 2), are selected for the classification.

The desirable features and thresholds are then put back to the class hierarchy to define the classifiers for each of the classes in eCognition. The classification with the features and thresholds calculated by this SEaTH algorithm does not produce fuzzy results, and each classified object has the full membership to its class. SEaTH algorithm calculates the features and thresholds based on the selected training objects, which should be representative enough to cover the entire spectral range of the class. Usually, the classification using the features and thresholds calculated by SEaTH do not classify all the objects due to that there are objects with feature values outside of the range of the selected training

objects. So, using SEaTH algorithm, there will often be a "back and forth" process. After select the training objects, calculate the features and thresholds, and classify the image, it is needed to observe which objects are not included (classified), and then reselect the training objects again, carefully include those objects, and recalculate the features and thresholds, and classify the image again. To avoid this back and forth process, based on the observed values of those not-classified objects, the thresholds of the calculated features can be adjusted in order to include those values, but it is recommended only when the missing objects are few, because adjusting the thresholds without recalculating can cause the mis-classification of other objects. All in all, the OBIA with features and thresholds identified with SEaTH algorithm produced a good classification result. With SEaTH, the time-consuming trial and error practice for seeking significant features and proper thresholds can be avoided. SEaTH algorithm can also help to minimize human involvement in classification steps and speeds up the process of classification when huge data is to be dealt with, however, it could be at the expense of accuracy.

Acknowledgement

Sincere thanks go to Professor Niemeyer Imgard and Prashanth Marpu for their kind help in introducing SEaTH algorithm and the help with the application of it.

References

Baatz, M., and Schaepe, A., 1999. Object-oriented and multi-scale image analysis in semantic networks. 2nd International Symposium: Operationalization of Remote Sensing, 16-20 August, ITC, the Netherlands.

Baatz, M., Heyen, M., and Hofman, P., 2004. eCognition professional, user guide 4. User guide. Definiens AG. http://www.definiens-imaging.com

Baatz, M., and Schaepe, A., 2000. Multiresolution segmentation optimization approach for high quality multi-scale image segmentation. In: J. Strobl, T. Blaschke and G. Griesebner (eds), Angewandte Geographische Informations-Verarbeitung XII, Wichmann, Karlsruhe, pp. 12-23.

Bhattacharyya, A., 1943, On a measure of divergence between two statistical

populations defined by their probability distributions, Bulletin of the Calcutta Mathematical Society 35, pp. 99-109.

Definiens, 2006. Definiens professional User Guide 5. Definiens AG, Munich.

Espindola, G. M., Camara, G., Reis, I. A., Bins, L. S., and Monteiro, A. M., 2006. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. International Journal of Remote Sensing, 27, pp. 3035-3040.

Franklin, J., Woodcock, C. E., and Warbington, R., 2000. Multi-Attribute Vegetation Maps of Forest Services Lands in California Supporting Resource Management Decisions. Photogrammetric Engineering & Remote Sensing, 66, 1209-1217.

Fukunaga, K., 1990, Introduction to Statistical Pattern Recognition. 2nd ed. Academic Press, London.

Hay, G.L., Gastilla, G., Wulder, M., Ruiz, J.R., 2005, An automated object-based approach for the multi-scale image segmentation of forest scenes. International Journal of Applied Earth Observation and Geoinformation 7, pp. 339-359.

Marpu P.R., Niemeyer I., Nussbaum S., Gloaguen R., 2006. A procedure for automatic object-based classification. In: Proc. 1st International Conference on Object-based Image Analysis (OBIA 2006), Salzburg, 4-5 July 2006, ISPRS Volume No. XXXVI - 4/C42.

Nussbaum, S., Niemeyer, I., & Canty, M. J., 2005. Feature Rcognition in the Context of Automated Object-Oriented Analysis of Remote Sensing Data Monitoring the Iranian Nuclear Sites. In: Proc. SPIE's Europe Symposium Optics/Photonics in Security & Defense, Bruges, 26-29 September 2005, SPIE Vol. ED103 (CD-Rom).

Nussbaum S., Niemeyer I., Canty M. L., 2006. SEaTH-A new tool for automated feature extraction in the context of object-based image analysis. In: Proc. 1st International Conference on Object-based Image Analysis (OBIA 2006), Salzburg, 4-5 July 2006, ISPRS Volume No. XXXVI – 4/C42.

Shackelford A. K., and Davis C. H., 2003. A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. IEEE Transactions on Geoscience and Remote Sensing, 41, pp. 2354-2364.

Tahir, M. A., and J. Smith, 2006, Improving Nearest Neighbour Classifier Using Tabu Search and Enhance Distance Metrics. Icdm, Sixth IEEE International Conference on Data Mining (ICDM' 06), pp. 1086-1090.

CHAPTER 6

MODIS EVI AS AN ANCILLARY DATA FOR AN OBJECT-BASED IMAGE ANALYSIS WITH MULTI-SPECTRAL MODIS DATA*

Abstract. This paper investigates the contribution of Enhanced Vegetation Index (EVI) data to the improvement of object-based image analysis using multi-spectral Moderate Resolution Imaging Spectral-radiometer (MODIS) imagery. Object-based image analysis classifies objects instead of single pixels. The idea to classify objects stems from the fact that most often the important information to process an image is not presented in single pixels but in groups of pixels (objects) (Blaschke *et al*. 2001). Based on image segmentation, object-based image analysis uses not only spectral related information, but spatial, textural and contextual information as well. However, which type of information to use depends on the image data and the application, among many other factors. Enhanced Vegetation Index (EVI) data are from the MODIS sensor aboard Terra spacecraft. EVI improves upon the quality of Normalized Difference Vegetation Index (NDVI) product. It corrects for some distortions in the reflected light caused by the particles in the air as well as the ground cover below the vegetation. The EVI data product also does not become saturated as easily as NDVI when viewing rainforests and other area of the Earth with large amounts of chlorophyll. In this research, 69 EVI data (scenes)

* Based on the content of this chapter, a paper titled "MODIS EVI as an ancillary data for an object-based image analysis with multi-spectral MODIS data" was presented in the conference GEOBIA 2008, 6-7, August, 2008, Calgary, Canada.

collected during the period of three years (from January of 2001 to December of 2003) in a mountainous vegetated area were used to study the correlation between EVI and the typical green vegetation growth stages. These data sets can also be used to study the phenology of the land cover types. Different land cover types show distinct fluctuations over time in EVI values and this information might be used to improve land cover classification of this area. Object-based image analysis was used to perform the land cover classification: one was only with MODIS multispectral data (seven bands), and the other one included also the 69 EVI images. Eight land cover types were distinguished and they are temperate forest, tropical dry forest, grassland, irrigated agriculture, rain-fed agriculture, orchards, lava flows and human settlement. The two classifications were evaluated with independent (from the training data) verification data, and the results showed that with EVI data, the classification accuracy was significantly improved, at 0.01% level, evaluated by the test of McNemar.

Key words: Enhanced Vegetation Index (EVI), Moderate Resolution Imaging Spectral-radiometer (MODIS), phenology, object-based image analysis

6.1. Introduction

In the last three decades, advances in computer technology, earth observation sensors and GIS science, led to the development of "Object-based image analysis" as an alternative to the traditional pixel-based image analysis. Many studies have shown that traditional pixel-based image analysis is limited because basically it uses only spectral information of single pixels and thus produces poor results especially with high spatial resolution satellite images. By contrast, object-based image analysis (OBIA) works on (homogeneous) objects. As an object is a group of pixels, its characteristics such as the mean, standard deviation of spectral values, etc. can be calculated and used in the classification; besides shape and texture features of the objects can also be derived and used to differentiate land cover classes with similar spectral information.

Enhanced Vegetation Index (EVI) data were obtained from the MODIS sensor

aboard Terra spacecraft. EVI improves upon the quality of NDVI product. It corrects for some distortions in the reflected light caused by the particles in the air as well as the ground cover below the vegetation. The EVI data product does not become saturated as easily as NDVI when viewing rainforests and other area of the Earth with large amounts of chlorophyll. The EVI data are designed to provide consistent, spatial and temporal comparisons of vegetation conditions, and it offers the potential for regional analysis and systematic and effective monitoring of the forest area. This paper investigates the contribution of EVI data to the improvement of the performance of OBIA with MODIS imagery. The 69 EVI data collected during the period 2001 to 2003 were plotted to obtain the phenology information of the land covers in the study area.

Phonology is the study of the times of recurring natural phenomena. One of the most successful of these approaches is based on tracking the temporal change of a Vegetation Index such as NDVI or EVI. The evolution of vegetation index exhibits a strong correlation with the typical green vegetation growth stages. The results (temporal curves) can be analyzed to obtain useful information such as the start/end of vegetation growing season. However, remote sensing based phonological analysis results are only an approximation of the true biological growth stages. This is mainly due to the limitation of current space based remote sensing, especially the spatial resolution, and the nature of vegetation index. A pixel in an image does not contain a pure target but a mixture of whatever intersected the sensor's field of view.

Since EVI is a good indicator of the phenology of the land cover types, the research tested the contribution of EVI data to the land cover classification. OBIA was used to classify the land covers in the study area: one was with single date MODIS multi-spectral reflectance data (seven bands), and the other one included also the EVI data (69 dates). The classification results were evaluated with independent verification data and were compared in order to examine the contribution of the EVI data to image classification using object based method with MODIS multi-spectral data.

6.2. Study are and Data

6.2.1 The study area

The study area is located in Michoacán state, central west of México, covering an area of approximately 58*60 km^2, within the longitude of 102° 00' W and 102° 32' W, and latitude of 19° 02' N and 19° 36' N (figure 6-1). The main land cover types in the study area are temperate forest, tropical dry forest, orchards, grassland, irrigated agriculture, rain fed agriculture, lava flow, and human settlement.

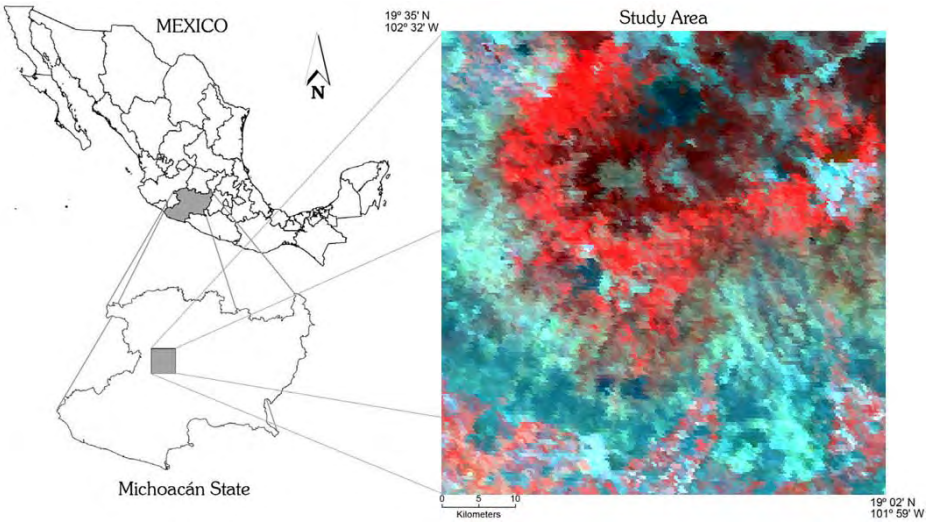


Figure 6-1: *The study area. Left side of the figure are two sketch maps indicating México and Michoacán state where the study area is located; right side is the false colour composite of MODIS imagery with R, G, B represented by bands 2 (near-infrared), band 1 (red), and band 3 (green).*

6.2.2 Earth Observation data

Data used in this research were acquired by the MODIS instrument on board Terra and Agua satellites from National Aeronautics and Space Administration (NASA). Terra MODIS and Agua MODIS take between one and two days to cover the entire Earth's surface, with a complete 16-day repeat cycle. Both

sensors acquire data in 36 spectral bands, or groups of wavelengths, and their spatial resolution (pixel size at nadir) is 250m for channels 1 and 2 (0.6 μm -0.9 μm), 500m for channels 3 to 7 (0.4 μm -2.1 μm) and 1000m for channels 8 to 36 (0.4 μm -14.4 μm). These channels are calibrated on orbit by a solar diffuser (SD) and a solar diffuser stability monitor (SDSM) system, which convert the Earth surface radiance to radio-metrically and geo-locationally calibrated products for each band (Xiong *et al.* 2003). Although recent evaluations have reported a geo-location error of 113m at nadir (Knight *et al.* 2006), official technical specifications warrant 50m geo-location accuracy (Wolfe *et al.* 2002). In both cases, the geo-location is less than half a pixel dimension, and hence acceptable for the multi-spectral analysis (Carrao *et al*. 2007).

The data acquired by the MODIS sensor is used to generate multiple products at different pre-processing stages. In this study the MOD09A1 product was used, a weekly composite of surface reflectance images, freely available from MODIS Data Product website (http://modis.gsfc.nasa.gov). This specific product is an estimate of the surface spectral reflectance imaged at a nominal spatial resolution of 500m for the first seven bands as it would have been measured at ground level if there were no atmospheric scattering or absorption. The applied correction scheme compensates for the effects of gaseous and aerosol scattering and absorption, for adjacency effects caused by variation of land cover, for Bidirectional Reflectance Distribution Function (BRDF), for coupling effects, and for contamination by thin cirrus (Vermote and Vermeulen 1999). Seven first spectral bands (1-7) of MOD09A1 imagery obtained on 08 March 2007 were used because they are closely related to land cover mapping. In addition, a set of 69 MODIS Enhanced Vegetation Index (EVI) images covering a full three years observation period, from January 2001 to December 2003 were also used for image classification.

EVI is an ‘optimized’ index designed to enhance the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmosphere influence. EVI is computed following this equation (Huete *et al*. 2002):

$$EVI = G * \frac{(NIR - \mathrm{Re}\, d)}{(NIR + c1 * \mathrm{Re}\, d - c2 * Blue + L)} \quad (6.1)$$

Where NIR/red/blue are atmospherically corrected or partially atmosphere corrected (Rayleigh and ozone absorption) surface reflectance, L is the canopy background adjustment that addresses non-linear, differential NIR and red radiant transfer through a canopy, and c1, c2 are the coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influence in the red band. The coefficients adopted in the MODIS EVI algorithm are: L = 1, c1 = 6, c2 = 7.5, and G (gain factor) = 2.5. Whereas the NDVI is chlorophyII sensitive, the EVI is more responsive to canopy structural variations, including leaf area index (LAI), canopy type, plant physiognomy, and canopy architecture. The two VIs complement each other in global vegetation studies and improve the detection of vegetation changes and extraction of canopy biophysical parameters. For this work, OBIA is carried out in an object-based image analysis software eCognition (Definiens 2006).

6.3. Methods

6.3.1 OBIA in eCognition

OBIA in eCognition comprises two parts: multi-resolution image segmentation and classification based on objects' features in spectral, spatial, and textural domains. Image segmentation is a kind of regionalization, which delineates objects according to a certain homogeneity criteria and at the same time requires spatial contingency (Lang *et al*. 2006). Several parameters are used here to guide the segmentation result. The scale parameter determines the maximum allowed heterogeneity for the resulting image objects. The colour criterion defines the weight with which the spectral values of the image layers contribute to image segmentation, as opposed to the weight of the shape criterion. The relationship between colour and shape criteria is: colour + shape = 1. Maximum colour criterion 1.0 results in objects spectrally homogeneous; while with a value of less than 0.1, the created objects would not be related to the spectral information at all. Smoothness is used to optimize image objects with regard to smooth borders, and compactness allows optimizing image

objects with regard to compactness (Baatz *et al*. 2004). The resulting objects also depend on the image data. For a given set of segmentation parameters, heterogeneous image data result in smaller image objects than homogeneous image data.

The image objects can then be classified either using a (standard) nearest neighbour (NN) classifier or fuzzy membership function, or a combination of both. The first classifier classifies the object primitives using user-defined sample objects, and by comparing the distance between object primitives and the training objects, the object primitives are classified into classes with which it has the shortest distance. The second classifier describes intervals of feature characteristics (Hofmann 2001b). The variety of object features can be used either to describe fuzzy membership functions, or to determine the feature space for NN. More detailed description of image segmentation and classification is given in Hofmann (2001a) and Gao *et al*. (2006). In this paper, the OBIA was performed with a standard NN classifier.

6.3.2. Accuracy assessment

Classification accuracy is used to describe the degree to which the derived image classification agrees with reality (Campbell 1996), and a classification error is, thus, the discrepancy between thematic map and reality. Accuracy assessment result is often represented by a confusion matrix. It is a simple cross-tabulation of the mapped class label against that observed one in the ground or reference data for a sample of cases at specified locations. The confusion matrix provides an obvious foundation for accuracy assessment (Campbell 1996, Canters 1997), by providing the basis to both describe classification accuracy and characterize errors, which may help to refine the classification. Many measures of classification accuracy can be derived from a confusion matrix (Stehman 1997). One of the most popular is the percentage of cases correctly allocated which is often regarded as *overall accuracy*, which is calculated by dividing the total number of the verification data with the number of the correctly classified image data. For the accuracy of individual classes, the percentage of cases correctly allocated may be derived from the confusion matrix by relating the number of cases correctly allocated to the class to the

total number of cases of that class. This can be achieved from two standpoints, depending on whether the calculations are based on the matrix's row (*user's accuracy*) or column marginal (*producer's accuracy*) (Foody 2002). The detailed description of the error matrix and the calculation of the measures of classification accuracy can be found in the second chapter 2.4.

Image classifications were evaluated with independent reference data which comprised of in total 496 points. These points were generated with a stratified random sampling method. Based on the land use map from the year 2000, random points were extracted from the 8 classes of interest. The properties of these points were interpreted based on the information from ortho-corrected photographs (1995), land use map from year 2000, and Landsat-7 ETM+ image (2003).

6.3.3. Test of McNemar

Map accuracy statements are often compared to evaluate the relative suitability of different classification techniques for mapping. Accuracy statements should be compared in a statistically rigorous fashion to provide a more objective basis for comment and interpretation (Foody 2004). In many remote sensing studies, the same set of ground data are used in the assessment of the accuracy of the thematic maps to be compared. For related samples, the statistical significance of the difference between two accuracy statements maybe evaluated using tests that take into account the lack of independence such as Test of McNemar. It is a non-parametric test that is based on confusion matrixes that are 2 by 2 in dimension. The attention is focused on the binary distinction between correct and incorrect class allocations. The McNemar test equation can be expressed as

$$Z^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \qquad (6.2)$$

In which f_{ij} indicates the frequency of verification samples lying in confusion matrix element i and j. f_{12} and f_{21} are the number of pixels that with one method were correctly classified, while with the other one were incorrectly

classified. Z^2 follows a chi-squared distribution with one degree of freedom (Agresti 1996). The statistical significance can be obtained with the derived z^2 compared against tabulated chi-square values. For example, with one degree of freedom, when the calculated z^2 >= 3.84, the two classifications are significantly different at 0.05% level; when z^2 >= 6.64, the two classifications are significantly different at 0.01% level; when z^2 >= 10.83, the two classification are significantly different at 0.001% level; and when z^2 < 3.84, the two classifications are not significantly different.

6.4. Results

6.4.1. Phenology of the land-cover types

69 EVI images from 2001 to 2003 were used to monitor the phenology of the eight land cover types.

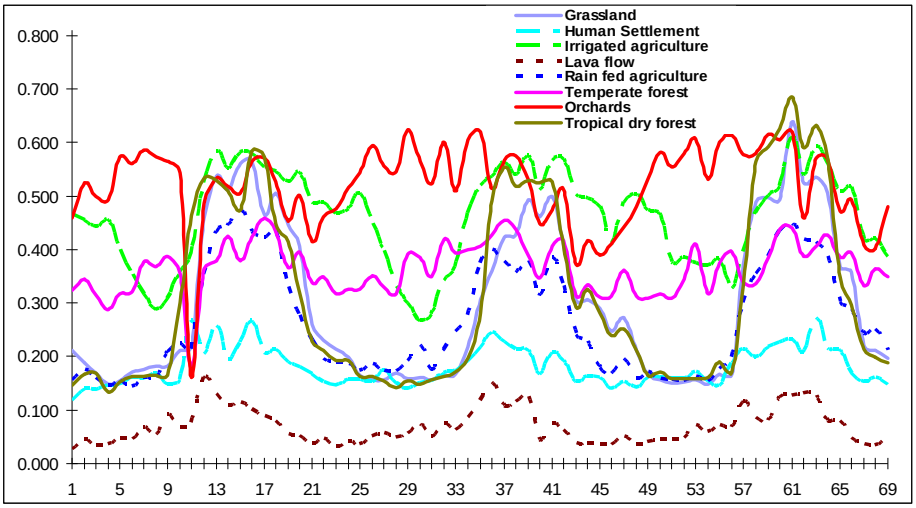


Figure 6-2: *EVI temporal profile of an area with 8 dominant land cover types over a period of 3 years (2001-2003), this temporal profile depicts the growing season every year as well as changes in this profile from year to year due to climate and other constraints.*

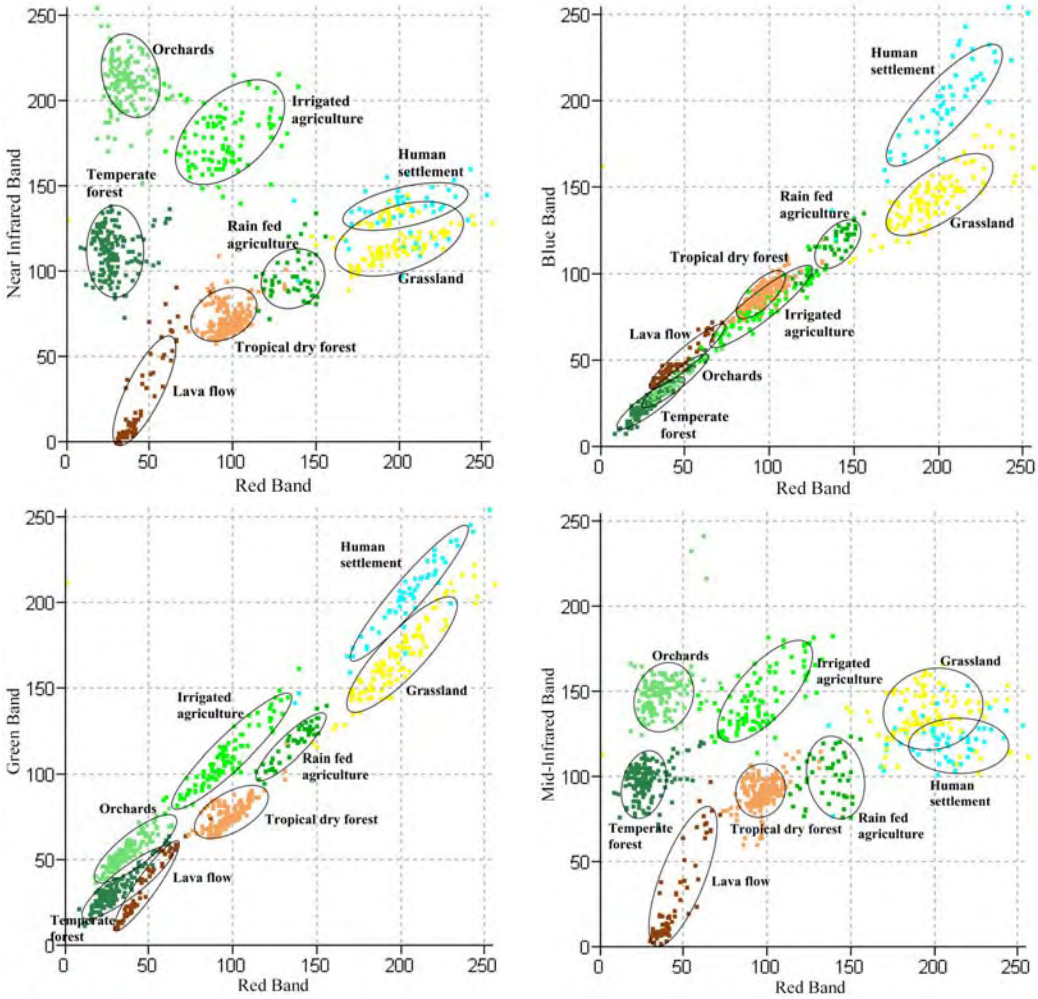
In figure 6-2, the horizontal axis represents the 69 MODIS EVI data in three years from which, each year there are 23 image data. The vertical axis represents the EVI values of the land cover types from different time of the

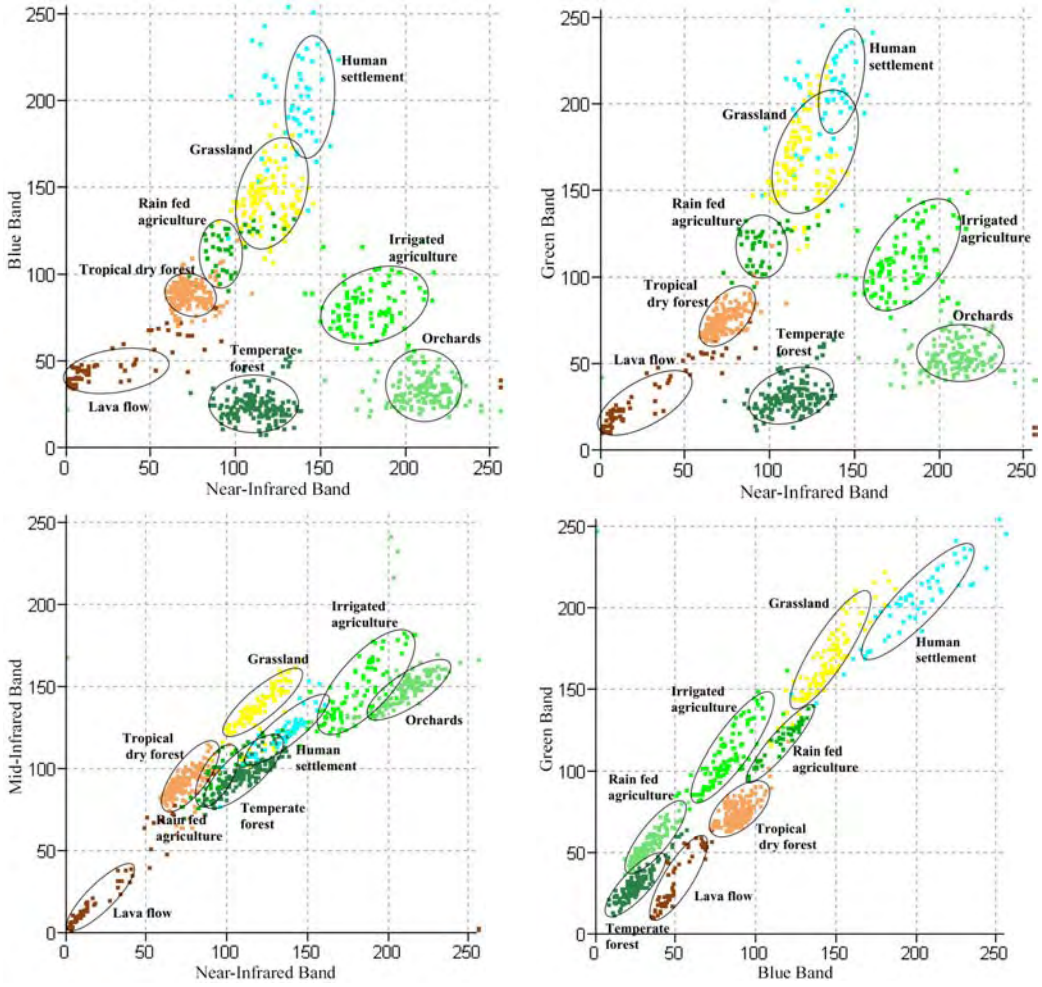
year. The 23 images from each year were taken every 16 days. In the example of the image data from 2001, roughly, image data 1 and 2 corresponds to data from January, 3 and 4 corresponds to February, and so on and so forth. The figure shows that for land cover types such as 'rain fed agriculture', 'tropical dry forest' and 'grassland', there are big changes in EVI values during different seasons in the period of three years; while for 'lava flow' and 'human settlement' the fluctuation of the EVI values during different seasons of the year is small. In the example of EVI values of 'tropical dry forest', the EVI values are low for image data 1 to 11, which corresponds in season to January – the end of May, dry season in México. In dry season, 'tropical dry forest' loses all the leaves and becomes totally dry, causing the rather low EVI values during this period. Its EVI values start to rise from data 11 and reach its peak on 17/18, which corresponds to the period from the end of May to the end of September and which is the rainy season in Mexico. Thus 'tropical dry forest' has high EVI values because in rainy season it regains its leaves and tends to have higher reflection in near infrared band. From data 18 its EVI value starts to drop and keeps dropping until data 28 which corresponds to March/April of the next year, during which the dry season starts and continues until the beginning of the rainy season of the next year. 'Tropical dry forest' and 'grassland' show evident seasonal change of EVI data because the growth of these two land cover types depends on the water from the natural rain. Other land cover types such as 'lava flow' and 'human settlement', having few vegetation cover, do not show much change in EVI values during different seasons of the year. Also, their EVI values are rather low among the eight land cover types (the EVI value of 'human settlement' is even lower than that of the 'tropical dry forest' during the dry season due to little vegetation cover). Figure 6-2 also shows that during the dry season, four land cover types 'grassland', 'human settlement', 'rain fed agriculture', and 'tropical dry forest' are not separable.

Land cover types such as 'orchard', 'irrigated agriculture', and 'temperate forest' show high EVI values during the year and small fluctuations during different seasons of the year. These land cover types keep green all the year round and show only small difference between dry season and rainy season. The sudden drop of the EVI value of data 11 for 'orchards' is due to the noise

of image data. In rainy season, the constant presence of rain clouds influences the image data quality and produce noises. One way to derive the missing value at this point is to average data value 10 and 12.

6.4.2 Separability analysis of the land-cover types




Blue Band
Near-Infrared Band
Human settlement
Grassland
Rain fed agriculture
Irrigated agriculture
Tropical dry forest
Temperate forest
Orchards
Lava flow
Green Band
Near-Infrared Band
Human settlement
Grassland
Rain fed agriculture
Irrigated agriculture
Tropical dry forest
Temperate forest
Orchards
Lava flow
Mid-Infrared Band
Near-Infrared Band
Irrigated agriculture
Grassland
Orchards
Tropical dry forest
Human settlement
Temperate forest
Rain fed agriculture
Lava flow
Green Band
Blue Band
Grassland
Human settlement
Irrigated agriculture
Rain fed agriculture
Rain fed agriculture
Tropical dry forest
Lava flow
Temperate forest

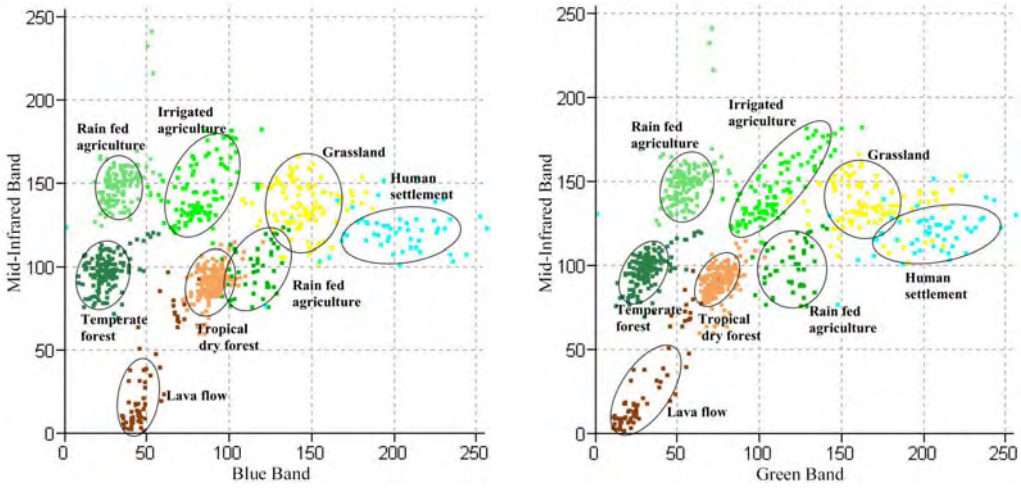


Figure 6-3: *Separability of eight land-cover classes in different combinations of the feature space.*

Figure 6-3 shows that land cover types are better separated when both near-infrared and optical spectral bands, or mid-infrared and optical spectral bands are present. In the feature spaces composed of only optical bands, or only near-infrared or mid-infrared bands, the separability between the land cover types appears to be low.

6.4.3 OBIA with single date multi-spectral MODIS imagery

First, the multispectral MODIS image was segmented using the following parameters: scale factor 20, colour 0.7, shape 0.3, and smoothness 0.5. By giving more weights to colour factor, the segmentation procedure considers more spectral information. For the shape factor, we gave the same weight to smoothness and compactness for the produced objects. The selection of the parameters was based on the visual checking. Altogether 3518 objects were created. In the class hierarchy, eight land cover classes were created and for each of them, the standard nearest neighbour classifier using the mean values of the spectral information from the 7 bands of MODIS multispectral image. Training samples were defined for each of the seven classes, and the image was classified and is presented in figure 6-4. The legend of the produced image is in figure 6-6.

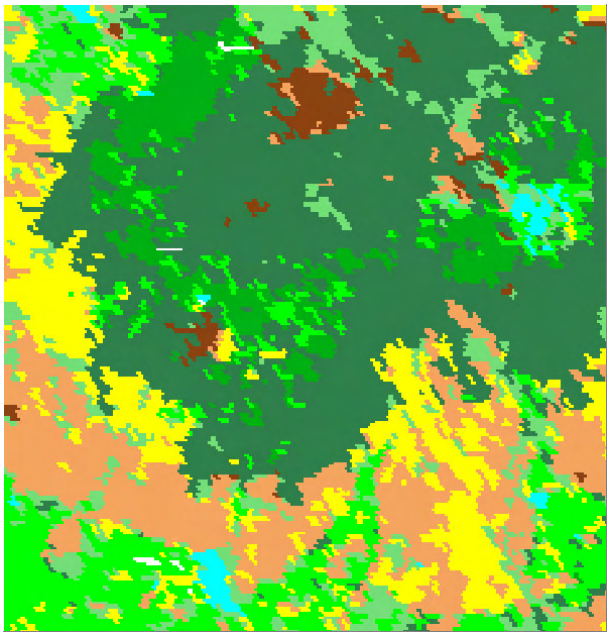


Figure 6-4: *Object-based image analysis with only reflectance data*

Figure 6-6: *Legend*

The classified image was evaluated with the independent verification data which is comprised of 499 stratified random sample points. Error matrix was produced and it is shown in table 6-1 (see annex). The classification obtained an overall accuracy 57.3%.

6.4.4 OBIA with both spectral reflectance data and multi-date EVI data

OBIA was performed with both single date spectral reflectance data and 69 dates EVI data. Image was segmented with only 7 spectral reflectance bands and with the same parameters in order to obtain the same number of objects as the first classification (section 6.4.3). The same class hierarchy and the same training samples were used for the eight classes of interest. NN classifier was also used here but based not only on the spectral information of the reflectance data, but also on the 69 EVI data. The image was classified and shown in figures 6-5. And the legend of the classified image was shown in figure 6-6.

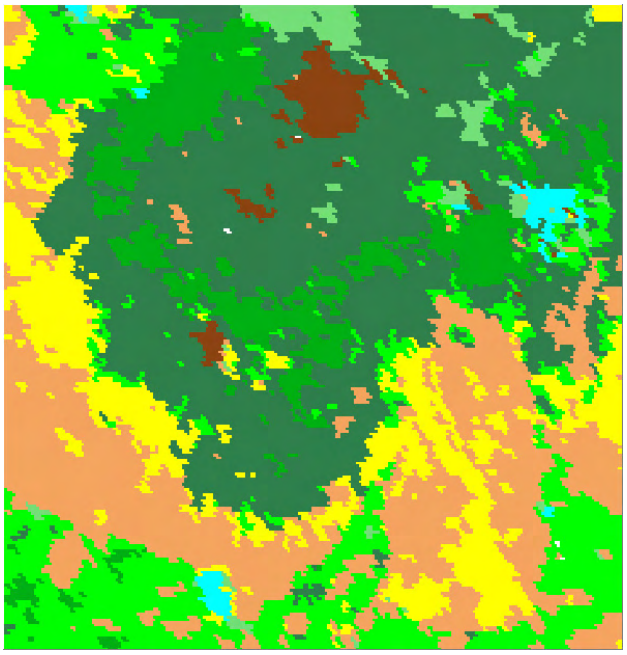


Figure 6-5: *Object-based image analysis with reflectance data and multi-date EVI data*

Figure 6-6: *Legend*

In the obtained result, land cover types were visually more homogeneous. The classified image was evaluated with the same independent reference data which is comprised of 499 stratified random sample points. Error matrix was produced and it is shown in table 6-2 (see annex). This classification obtained an overall accuracy 62.5%.

6.4.5 Comparison and discussion

The two classifications were compared. The OBIA using both single date spectral reflectance data and multiple dates EVI data shows an improvement of 5.2% in accuracy than that obtained by using only single date spectral reflectance data. Test of McNemar was used to evaluate the significance of the difference between these two classifications. 249 pixels were correctly classified and 150 pixels were wrongly classified by both classifications, 63 pixels were correctly classified by the classification with both spectral reflectance and EVI data, and 37 pixels were correctly classified by the

classification using only spectral reflectance data. The calculated chi-square z^2 = 6.76, Checked by the significance table, the two classifications are significantly different, at 0.01% level. The result shows that the OBIA with both spectral reflectance and EVI data obtained the accuracy significantly higher than the classification with only spectral reflectance data. Further checking these two classifications shows that the improvement is mainly represented by classes 'tropical dry forest', 'lava flow', 'orchards', and 'irrigated agriculture', which contributed 3.8%, 0.8%, 1.0%, and 2.2% to the improvement of the accuracy, respectively. As shown in the accuracy assessment tables, by adding the EVI data, the spectral confusion between 'tropical dry forest' and 'rain fed agriculture' in the MODIS image was alleviated and it improved 15.4% the produced accuracy of 'tropical dry forest'. EVI data alleviated the confusion between 'lava flow' and 'tropical dry forest'; between 'irrigated agriculture', and 'grassland'; and between 'orchards' and 'irrigated agriculture'.

The MODIS spectral reflectance data contain information from visible, near infrared and mid infrared spectral regions which record the most important spectral signature of the classes for land cover mapping. Due to the coarse spatial resolution of 250m and 500m, multi-spectral MODIS image was composed mostly of mixed pixels, which reduced the spectral separability between classes of interest. The multi-date EVI data helped to separate land cover classes which were difficult to be differentiated with only single date spectral reflectance information. For example the 'grassland' showed serious spectral confusion with 'rain fed agriculture' in the spectral reflectance image; also 'tropical dry forest' and 'lava flow'; and 'rain fed agriculture' and 'tropical dry forest'. By observing figure 6-2, we can see that these pairs of classes exhibit different reflectance values and are separable using EVI data.

6.5 Conclusions

This research investigates the contribution of EVI data to the improvement of OBIA with MODIS spectral reflectance data. The evolution of the 69 EVI data collected during the period 2001 to 2003 exhibits a strong correlation with the phenology of the land cover types. The results (temporal curves) were analyzed

and various land cover types show different fluctuations in EVI values, which were used to improve the land cover classification. OBIA was carried out for land cover classification: one was only with MODIS spectral reflectance data, and the other one included also the 69 EVI images. The classifications were evaluated with independent verification data, and the results showed that with EVI data, the accuracy was significantly improved, at a level of 0.01%, by the test of McNemar. This paper shows that the MODIS EVI data supply important information not only to monitor the phenology of the land cover types, but to differentiate land cover types which were difficult to be differentiated using single date spectral reflectance data.

Acknowledgement

The authors thank CONACYT-CONAFOR 2005-C02-14741 for supplying PhD fellowship to the first author during writing up of the paper.

References

Baatz, M., and Schaepe, A., 1999. Object-oriented and multi-scale image analysis in semantic networks. 2nd International Symposium: Operationalization of Remote Sensing, 16-20 August, ITC, the Netherlands.

Baatz, M., Benz, U., Dehghani, S., Heynen, M., Holtje, A., Hofmann, P., Lingenfelder, I., Mimler, M., Sohlbach, M., Weber, M., Willhauck, G., 2004, eCognition User's Guide. http://www2.definiens.com/central/default.asp (the last accessed 07/March/2006).

Bastin, L., 1997, Comparison of fuzzy c-means classification, linear mixture modeling and MLC probabilities as tools for unmixing coarse pixels. International Journal of Remote Sensing, 18, pp. 3629-3648.

Blaschke, T., and Strobl, J., 2001, What's wrong with pixels? Some recent development interfacing remote sensing and GIS. GeoBIT/GIS, 6, pp. 12-17.

Campbell, J.B., 1996, Introduction to Remote Sensing (2nd ed.). London: Taylor and Francis.

Canters, F., 1997, Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification. Photogrammetric Engineering and Remote Sensing, 63, pp. 403-414.

Carrao, H., Goncalves, P., Caetano, M., 2007, Contribution of multispectral and multitemporal information from MODIS images to land cover classification. Remote Sensing of Environment, doi: 10.1016/j.rse.2007.07.002.

Dean, A.M., Smith, G.M., 2003, An evaluation of per-parcel land cover mapping using maximum likelihood class probabilities. International Journal of Remote Sensing, 24, pp. 2905–2920.

Definiens, 2006. Definiens professional User Guide 5. Definiens AG, Munich.

Foody, G.M., 2002, Status of land cover classification accuracy assessment. Remote sensing of environment, 80, pp. 185-201.

Gao, Y., Mas, J. F., Maathuis, B. H. P., Zhang, X. M., and Van Dijk, P. M., 2006. Comparison of pixel-based and object-oriented image classification approaches-a case study in a coal fire area, Wuda, Inner Mongolia, China. International Journal of Remote Sensing, 27, pp. 4039-4051.

Hay, G.J., and Gastilla, G., 2006, Object-based image analysis: strength, weakness, opportunities, and threats (SWOT). 1st International Conference on Object-based image analysis (OBIA 2006), 4-5, July, 2006, Salzburg, Austria.

Hofmann, P. 2001a. Detecting buildings and roads from IKONOS data using additional elevation information. In: GeoBIT/GIS 6, pp. 28-33.

Hofmann, P., 2001b. Detecting urban features from IKONOS data using an object-oriented approach. In: Remote Sensing & Photogrammetry Society (Editor): Proceedings of the First Annual Conference of the Remote Sensing & Photogrammetry Society 2001, pp. 28-33.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002, Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, 83, pp. 195-213.

Integrated Land and Water Information System (ILWIS), 2005. International Institute for Geo-Information Science and Earth Observation (ITC). The Netherlands

Justice, C.O., Markham, B.L., Townshend, J.R.G., and Kennard, R.L., 1989,

Spatial degradation of satellite data. International Journal of Remote Sensing, 10, pp. 1539-1561.

Lillesand, M., Kiefer, W.R., 1994, Remote sensing and image interpretation, Third Edition, John Wiley & Sons, Inc., New York, 750 pp.

Smits, P.C., Dellepiane, S.G., Schowengerdt, R.A., 1999, Quality assessment of image classification algorithms for land-cover mapping: a review and proposal for a cost-based approach. International Journal of Remote Sensing, 20, pp.1461-1486.

Stehman, S.V., 1997, Selecting and interpreting measures of thematic classification accuracy. Remote sensing of Environment, 62, pp. 77-89.

Vermote, E.F., Vermeulen, A., 1999, atmospheric correction algorithm: Spectral reflectances (MOD09) algorithm technical background document. Available online at http://modis.gsfc.nasa.gov/data/atbd/atbd_mod08.pdf/ (accessed 16 December 2007).

Wolfe, R.E., Nishihama, M., Fleig, A.J., Kuyper, J.A., Roy, D.P., Storey, J.C., Patt, F.S., 2002, Achieving sub-pixel geolocation accuracy in support of MODIS land science. Remote Sensing of Environment, 83, pp. 31-49.

Xiong, X., Sun, J., Esposito, J., Guenther, B., & Barnes, W.L., 2003, MODIS reflective solar bands calibration algorithm and on-orbit performance. Proceedings of SPIE-Optical Remote Sensing of the Atmosphere and Clouds III, Vol. 4891, pp. 95-104.

CHAPTER 7

CONCLUSIONS

7.1. The contribution of the thesis

This research set out to investigate the contribution of image segmentation and object based image analysis (OBIA) to land-cover/use classification using remote sensing images in a diverse and complex (regarding to the distribution of land cover types) study area. Project introduction, motivation, research questions, objectives and sub-objectives are summarized in chapter 1. The introduction to the theoretic background of OBIA including image understanding, image segmentation, fuzzy classification, and accuracy assessment is presented in chapter 2. In chapter 3, 4, 5, and 6, several experiments focused mainly on the application of image segmentation and OBIA were carried out, in comparison to pixel-based image analysis.

7.1.1. The effects of image segmentation to the classification accuracy

Image segmentation is a kind of regionalization, which delineates units according to a given criterion of homogeneity, and, at the same time, requires spatial contingency (Lang and Blaschke 2006). Image segmentation is a preliminary and critical step in OBIA, the assumption being that results directly affecting the performance of the object based classification. Image segmentation evaluation before the classification ensures that the best segmentation result is used in the classification stage. One principal point of

concern here is the selection of segmentation parameters, which has conventionally been based on trial-and-error approaches. In this part of the research, image segmentation was evaluated with an objective function proposed by Espíndola *et al*. (2006).

Landsat-7 ETM+ imagery was segmented with nine different parameter settings in the image processing software SPRING (from INPE, Brazil), which uses two parameters-a spectral threshold and a spatial/size limit-to guide the segmentation results. The segmented images were evaluated by the objective function (Espindola *et al*. 2006) which takes into account of the objects' intrasegment homogeneity and intersegment separability. The segmented images were then classified and the classification results were evaluated with independent ground data. It is shown that the best segmentations, in terms of objective function rating, also led to the classifications with the highest accuracies, and, the accuracy values presented similar distribution as the objective function values in the function of the segmentations. To be comparable, pixel based classification using the same set of training data was carried out and its accuracy was evaluated. The test of McNemar shows that the difference in accuracies between the pixel-based and OBIA is significant. This research shows that the objective function is an effective way to choose the optimal segmentation to carry out the classification; OBIA based on the optimal segmentation is superior than the pixel based one and it obtained accuracy which is significantly higher than that of the pixel based one.

7.1.2. OBIA with different spatial resolution images.

The performance of pixel-based and object-based image analysis was compared over satellite images with different spatial resolutions. The coarser spatial resolution images were generated by degrading spatial resolution of a multi-spectral SPOT-5 imagery which has a 10 m spatial resolution. Two simulating methods were used: 1) "mean-filtering" which averages the pixel values in a certain sized window (3*3, 10*10, 25*25), and 2) "cubic-filtering", modified from Justice *et al*. (1998), which keeps the local contrast in the image during the simulation procedure and produced images closer to the "real" images. These two sets of images were classified by both pixel-based maximum

likelihood classifier (MLC) and minimum distance classifier, and by object-based Nearest Neighbour (NN) classifier, respectively. Accuracy assessment results showed that for SPOT-5 and a simulated image with 30 m spatial resolution generated by cubic-filtering, OBIA obtained higher accuracies than those produced by pixel-based MLC and minimum distance classifiers. While, with images of further increased spatial resolutions (100m, and 250m), OBIA did not obtain higher overall accuracies than the pixel-based ones. This study showed that OBIA has advantage over the pixel-based one, but in the rating of the obtained accuracies, the advantage only holds true for higher spatial resolution images.

Comparing with pixel-based image classification methods, OBIA has many advantages: the way it classifies an image by partitioning it into objects is similar to the way humans comprehend the landscape; image-objects exhibit useful features (shape, texture, context relations with other objects) that single pixels lack; image-objects can be more readily integrated in vector GIS, etc. Based on this work, a general conclusion was drawn that the advantage of OBIA over the pixel-based one, based on classification accuracy rating, was only represented by images with higher spatial resolutions (< 30m). This is because that increased spectral variability within high spatial resolution imagery confuses traditional pixel-based classifiers, while by OBIA, pixels with similar spectral information are firstly grouped into objects and then classified. This way simplifies the image understanding and minimizes speckled effects in the classified images. Images with medium to low spatial resolution have lower spectral variability and thus are easily handled by pixel-based method. As for the images with low spatial resolution, such as MODIS, by applying OBIA, pixels belonging to different land cover types could possibly be grouped together thus are miss-classified and produce lower accuracy than that by pixel-based method. However, as stated before, the obtained accuracy values from this part of the research should never be taken in an absolute sense; there are too many factors which influence classification accuracy such as image data quality, the reliability of training data and testing/verification data, the accuracy assessment method adopted, among others.

As to the accuracy assessment method, though accuracy evaluated by random points (simple/stratified random) is generally recognized to be more trustworthy than that evaluated using homogenous areas of land cover types, it may not hold true when it comes to classification results generated by OBIA. Due to that single points/pixels are usually merged into surroundings by object based method, and thus, evaluated by single pixels, the accuracy of object-based image classification could be under-evaluated, which may explain that although the object-based classification results appear to be more appealing, their accuracy results evaluated with random points tend to be low in the experiments.

7.1.3. OBIA with a semi-automatic feature extraction algorithm.

Image objects can be characterized by a variety of spectral, spatial, texture, and contextual features. A comprehensive feature analysis is essential to work with image objects and the process of looking for the proper features could hinder the proper utilization of the strength of image objects. Separability and Threshold (SEaTH) algorithm is able to evaluate statistically any number of given features for any number of object classes of interest. It utilizes the advantage of a wide feature basis of objects and provides the fundament for a successful classification. With the SEaTH algorithm, the optimum features which characterize a class can be identified to separate the objects of interest. The working process of SEaTH algorithm can be generally described as the following: first, select training objects for the classes of interest (down to the sub-classes level) which should comprise 2-5% of the created objects; then export the feature values of the training objects and SEaTH algorithm calculates the desirable features and thresholds based on the separability. For each class, SEaTH compares it to the other classes using each of the selected features. The features and thresholds were listed according to the separability values. Based on the calculation, features and thresholds with the highest separability-the highest Jeffries-Matusita (J) value, which is in the range of (0, 2), are selected for the classification.

The desirable features and thresholds are then put back to the class hierarchy to define each of the classes in eCognition. The classification with the features

and thresholds calculated by this SEaTH algorithm does not produce fuzzy results, and each classified object has the full membership to its class. SEaTH algorithm calculates the features and thresholds based on the selected training objects, which should be representative enough to cover the entire spectral range of the class. Usually, the classification using the features and thresholds calculated by SEaTH does not classify all the objects due to that there are objects with feature values outside of the range of the selected training objects. So, using SEaTH algorithm, there will often be a "back and forth" process. After selecting the training objects, calculating the features and thresholds, and classifying the image, it is needed to observe which objects are not included (classified), and then reselect the training objects again, including those objects, and recalculate the features and thresholds, and classify the image again. To avoid this back and forth process, based on the observed values of those not-classified objects, the thresholds of the calculated features can be adjusted in order to include those values, but it is recommended only when the missing objects are few, because adjusting the thresholds without recalculating can cause the mis-classification of other objects. All in all, the OBIA with features and thresholds identified with SEaTH algorithm produced a good classification result. With SEaTH, the time-consuming trial and error practice for seeking significant features and proper thresholds can be avoided. SEaTH algorithm can also help to minimize human involvement in classification steps and speeds up the process of classification when huge data is to be dealt with, however, it could be at the expense of accuracy.

7.1.4. MODIS EVI as an ancillary data for an object-based image analysis with multi-spectral single date MODIS imagery

This research section works for a project which involves the use of MODIS imagery by investigating the contribution of Enhanced Vegetation Index (EVI) data to the improvement of OBIA using Moderate Resolution Imaging Spectral-radiometer (MODIS) imagery. Experiment shows that the evolution of the 69 EVI data collected during the period 2001 to 2003 (23 images for each year) exhibits a strong correlation with the typical green vegetation growth stages and various land cover types show different fluctuation in EVI values over time. This information was used to improve the land cover classification

with OBIA. Two experiments with OBIA were carried out: one was only with MODIS spectral reflectance data, and the other one included also the 69 EVI images. The classifications were evaluated with independent verification data, and the results showed that with EVI included in the classification, the accuracy was significantly improved, at a level of 0.01%, by the test of McNemar. It shows that the multi-temporal MODIS EVI data supply important information not only to monitor the phenology of the land cover types, but to differentiate land cover types which were difficult to be distinguished using single date spectral reflectance data.

7. 2. Further studies

7.2.1. Image segmentation parameterization.

Segmentation is not new, but only a few of the existing approaches are widely available in commercial software packages and lead to qualitatively convincing results while being robust and operational (Blaschke 2004). One reason is that the segmentation of an image into a given number of regions is a problem with a huge number of possible solutions, and experiments show that image segmentation has no unique solution (Lang *et al*. 2006). For example, segmentation results differ by changing only the bit depth (8 bit, 11 bit, etc.) of the image heterogeneity measure. The high degrees of possibility must be reduced to a few which are satisfying the given requirements.

Additionally, segmentation needs to address a certain scale, and to define a certain spectral and spatial factor. Up to now, the parameterization of image segmentation step is still based on trail and error approach. There is no an absolute measure to tell which parameter settings are especially suited for the analysis.

7.2.2. Accuracy assessment of OBIA

When performing the comparison between the pixel-based and object-based image classifications, in order to make sure the comparison is fare, the same set of reference data is often used (as was done in this research). Image classification accuracy evaluated by random points (simple random or stratified

random points) is generally recognized to be more trustworthy than that evaluated using homogenous areas of land cover types. However, this research found that the method of using random sampling points does not seem to suit to evaluate the results produced by OBIA (chapter 4 of this research). Also in Blaschke (2004), it observed that "for the real accuracy assessment - the comparison with ground truth, still many problems remain. The comparison with the landcover data sets can not be based on pixel statistics any more…"

Due to that single points/pixels are usually merged into surroundings to form objects by image segmentation, and thus, evaluated by single pixels, object based image analysis results could be under-evaluated. One way to improve this point is to generate the ground data not only based on the ground survey/interpretation of single pixels, but taking into account the surrounding pixels as well.

Generally speaking, the evaluation methods of object based image analysis results have not yet been well established. Technically, polygons produced by image segmentation/object based image analysis have to be compared with polygons of the verification data through various GIS overlay techniques taking into account the varying sizes and different degrees of overlap and misclassification, respectively (Blaschke 2004). And thus it requires the adaptation of the existing accuracy assessment methods and developing proper techniques that explicitly assess the accuracy of object-specific features.

7.3. General conclusions

Object-based image analysis (OBIA) is seen as an emerging discipline (Hay and Castilla 2006) or a new paradigm in image analysis (Lang and Blaschke 2006). The main reasons that drive the development of OBIA are (Hay *et al*. 2005): 1) A dramatic increase in the number of available high resolution remote sensing imagery. 2) An ever-growing user needs and expectations of geography information products. 3) Recognition of limitations with pixel-based image classification approaches (i.e. pixels are not true geographical objects and pixel topology is limited; current remote sensing image analysis largely neglects the spatial photo-interpretive elements - texture, context, shape;

increased spectral variability within high resolution imagery confuses traditional pixel-based classifiers and results in lower classification accuracies). 4) Awareness that object-based methods can make better use of neglected spatial information within remote sensing images, and provide greater integration with GIS information.

OBIA partitions imagery into (meaningful) image objects, and assesses the objects' spatial, spectral, and temporal characteristics. The core concept of OBIA is that the important information necessary to interpret an image is not represented in single pixels, but in meaningful image objects and their mutual relationships. Usually image objects are created by image segmentation. Existing segmentation techniques vary (a) in their approaches by starting at the entire image which is then subdivided, or starting with the individual pixels which are then grouped, and (b) in the weight given to the spatial component compared to the variables from the feature space (measurement space). In general, the more accurate results are obtained with the computationally more intensive techniques, i.e., starting with individual pixels and careful selection of the merges to be made. Segmentation studies have been performed for several decades, but only recently it is implemented with a high level of accuracy and fast performance due to the increased computer processing capacity. This resulted in the development and rapid expansion of OBIA. Based on image segmentation, object-based image analysis is a sub-discipline of GIS science devoted to partitioning remote sensing imagery into meaningful image-objects, and assessing their characteristics through spatial, spectral and temporal scales.

Instead of analyzing the spectral behavior of individual pixels, neighbouring pixels are grouped into objects, which then form the observation units. This grouping overcomes the problem of artificial square objects as used in per-pixel analysis, as long as the objects of interest cover a number of pixels allowing a meaningful representation of their shape. In contrast to pixel-based procedure, image objects can carry many more attributes than only spectral information. When considering image objects instead of pixels, a wealth of additional features can be used for characterization. Both statistically aggregated spectral features and geometrical and neighbourhood-related features can be used, by which the dimensionality of the resulting feature space

is getting significantly higher and rises to virtually limitless extent (Lang *et al*. 2006). If still a network of image objects is created, vertical relationships between super-objects and sub-objects in the object hierarchy can be used to expand the feature space even further (Lang 2005, Lang *et al*. 2006). Regarding to the advantages of OBIA, it can be generalized as (Hay and Castilla 2006): 1) it analyzes imagery by first partitioning an image into objects and then analyzes them. This is similar to the way humans conceptually organize the landscape to comprehend it. 2) Image-objects exhibit useful features, for example shape, texture, context relations with other objects that single pixels lack. 3) Image-objects can be more readily integrated in vector GIS than pixel-wise classified raster maps. 4) It uses image-object as basic processing units which reduces computational classifier load and enables the user to take advantage of non-parametric classifier.

Object based image analysis supports image understanding. Undoubtedly, visual interpreters can benefit from object based image analysis, which "offers valuable methodological assets in breaking down scene complexity into meaningful image primitives" (Lang and Blaschke 2006), through which an image can be analyzed in adaptive scales according to the domain of interest (Lang and Blaschke 2006).

Depending upon the scales at which the objects are assessed, humans recognize discrete objects, their size, shape, spatial arrangement and context changes (Marceau 1999). Performing object based image analysis, a clear understanding of the target scale, the target object set, and the target class scheme is a prerequisite. When it is clear how target objects are structured by sub-level primitives, multi-scale segmentation can help to reduce the complexity in images through aggregating information by scaled representations. The appropriate segmentation levels helps to match the solid image with the abstract reality (Lang 2005). A skilled interpreter who has a better understanding of the above mentioned prerequisite can master better the available object features and their mutual relationships in looking for the interested features. Image is then analyzed by putting expert knowledge into a rule system which categorizes the image objects by their spectral and spatial properties, and their mutual relationships (Bruce & Green 1990).

As the last paragraph of the conclusion part, I would like to add a few more words to the contribution of object based image analysis technique to the application of land cover mapping in México. As is well known, land cover mapping in Mexico has been based on visual interpretation in which interpreters utilize the colour, shape, texture information, etc. to deliver the mapping results. Visual interpretation of aerial photos or satellite imagery can produce detailed land-cover thematic maps with better accuracy than digital image classification (Sader and Winne 1992, Mas and Ramirez 1996). However, the method is slow, expensive (in terms of time expenditure and expertise requirements) and to some extent subjective. Object based image analysis also utilizes the colour, shape, texture information, etc, which tries to mimic the way humans interprets the image. Image segmentation delineates image objects based on colour and shape criteria, which is much fast than the manual digitization. At this point, object based image analysis can contribute to land cover mapping in Mexico in a positive way. Though, since it is automatic digital image classification, the object-based approach has its limitations and the human perception is still an ultimate benchmark, undefeated in analyzing complex scene contents with ease.

References

Blaschke, T., 2004, Object-based contextual image classification built on image segmentation. IEEE, 0-7803-8350—8/04 (C), pp. 113-119.

Bruce, V. & P.R. Green, 1990. Visual Perception. Physiology, Psychology and Ecology. Lawrence Erlbaum Associates, East Sussex.

Hay, G.J., and Gastilla, G., 2006, Object-based image analysis: strength, weakness, opportunities, and threats (SWOT). 1st International Conference on Object-based image analysis (OBIA 2006), 4-5, July, 2006, Salzburg, Austria.

Lang, S., and Blaschke, T., 2006, Bridging remote sensing and GIS - what are the main supportive pillars? Object-based image analysis: strength, weakness, opportunities, and threats (SWOT). 1st International Conference on Object-based image analysis (OBIA 2006), 4-5, July, 2006, Salzburg, Austria.

Lang, S., 2005, Image object vs. landscape objects. Interpretation, hierarchical

representation and significance. Unpublished PhD thesis, Salzburg University.

Marceau, D., 1999. The scale issue in the social and natural sciences. Canadian Journal of Remote Sensing 25, 347-356.

Openshaw, S., 1984, The modifiable areal unit problem, In: Concepts and Techniques in Modern Geography (CATMOG), No, 38, 40p. GeoBooks, Norwich, United Kingdom.

ANNEX

	Reference data									
Classification data*	Grassland	Human settlement	Irrigated agriculture	Lava flow	Rain fed agriculture	Temperate forest	Orchards	Tropical dry forest	Total	User's Accuracy (%)
Grassland	21	2	2	0	2	2	0	6	35	60
Human settlement	0	46	1	0	0	0	0	1	48	95.8
Irrigated agriculture	1	2	32	0	0	0	0	0	35	91.4
Lava flow	0	0	0	45	0	0	0	0	45	100
Rain-fed agriculture	1	0	2	0	30	2	1	4	40	75.0
Temperate forest	4	6	6	9	6	152	6	8	197	77.2
Orchards	0	2	1	0	0	3	34	0	40	85.0
Tropical dry forest	5	1	14	0	3	24	1	112	160	70.0
Total	32	59	58	54	41	183	42	131	600	
Producer's Accuracy (%)	65.6	78	55.2	83.3	73.2	83.1	81	86		
Overall accuracy = (21+46+32+45+30+152+34+112) / 600 * 100% = 78.7%										

Table 5-1: Accuracy assessment for the classification produced by SEaTH algorithm.

	Reference data									
Classification data*	Grassland	Human settlement	Irrigated agriculture	Lava flow	Rain fed agriculture	Temperate forest	Orchards	Tropical dry forest	Total	User's Accuracy (%)
Grassland	17	2	9	0	6	18	1	34	87	19.5
Human settlement	0	22	0	0	0	0	1	0	23	95.7
Irrigated agriculture	1	6	36	0	3	3	10	1	60	60
Lava flow	1	0	0	19	0	3	0	0	23	82.6
Rain-fed agriculture	4	9	7	0	20	10	1	14	65	30.8
Temperate forest	4	1	3	2	4	96	8	12	130	73.9
Orchards	0	0	0	0	1	7	14	0	22	63.6
Tropical dry forest	3	0	0	6	3	15	0	62	89	69.7
Total	30	40	55	27	37	152	35	123	499	
Producer's Accuracy (%)	56.7	55	65.5	70.4	54.1	63.2	40	50.4		
Overall accuracy = (17+22+36+19+20+96+14+62) / 600 * 100% = 57.3%										

Table 6-1: Accuracy assessment for the classification produced object based image analysis using only spectral reflectance value.

	Reference data									
Classification data*	Grassland	Human settlement	Irrigated agriculture	Lava flow	Rain fed agriculture	Temperate forest	Orchards	Tropical dry forest	Total	User's Accuracy (%)
Grassland	17	1	1	0	7	24	0	27	77	22.1
Human settlement	0	20	0	0	0	0	1	0	21	95.2
Irrigated agriculture	2	9	47	0	4	5	4	5	76	61.8
Lava flow	2	1	0	23	3	5	0	0	34	67.7
Rain-fed agriculture	0	6	1	0	13	4	0	0	24	54.2
Temperate forest	2	1	0	3	7	92	10	10	125	73.6
Orchards	0	1	0	0	1	6	19	0	27	70.4
Tropical dry forest	7	1	6	1	2	16	1	81	115	70.4
Total	30	40	55	27	37	152	35	123	499	
Producer's Accuracy (%)	56.7	50	85.5	85.2	35.1	60.5	54.3	65.9		
Overall accuracy = (17+20+47+23+13+92+19+81) / 600 * 100% = 62.5%										

Table 6-2: Accuracy assessment for the classification produced object based image analysis using both spectral reflectance data and EVI data.