URBAN LAND COVER CLASSIFICATION AND MOVING VEHICLE EXTRACTION USING VERY HIGH RESOLUTION SATELLITE IMAGERY

BAHRAM SALEHI

September 2012

TECHNICAL REPORT
NO. 281

URBAN LAND COVER CLASSIFICATION AND MOVING VEHICLE EXTRACTION USING VERY HIGH RESOLUTION SATELLITE IMAGERY

Bahram Salehi

Department of Geodesy and Geomatics Engineering
University of New Brunswick
P.O. Box 4400
Fredericton, N.B.
Canada
E3B 5A3

September, 2012

PREFACE

This technical report is a reproduction of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Geodesy and Geomatics Engineering, September 2012. The research was supervised by Dr. Yun Zhang, and funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) Strategic Project Grants and Discovery Grants Programs.

As with any copyrighted material, permission to reprint or quote extensively from this report must be received from the author. The citation to this work should appear as follows:

Salehi, Bahram (2012). *Urban Land Cover Classification and Moving Vehicle Extraction Using Very High Resolution Satellite Imagery*. Ph.D. dissertation, Department of Geodesy and Geomatics Engineering, Technical Report No. 281, University of New Brunswick, Fredericton, New Brunswick, Canada, 182 pp.

ABSTRACT

This Ph.D. dissertation reviews the current techniques and develops improved techniques for the analysis of very high resolution (VHR) imagery of urban areas for two important applications: land cover classification and moving vehicle (and velocity) extraction.

First, a comprehensive review is conducted on the current literature in the area of urban land cover classification of VHR imagery. The review discusses the usefulness of two groups of spatial information used in both pixel-based and object-based classification approaches. The first group is spatial information inherent in the image such as textural, contextual, and morphological (e.g., shape and size) properties of neighboring pixels, and the second group is the spatial information derived from ancillary data such as LiDAR and GIS vector data. The review provides guidelines on the use of spatial information for urban land cover classification of VHR images.

Second, a novel multisource object-based classification framework is developed using the Cognition Network Language available in the eCognition® software package. The framework integrates VHR images and height point data for detailed classification of urban environments. The framework addresses two important limitations of the current literature: the transferability of the framework to different areas and different VHR images, and the impact of misregistration between different data layers on classification accuracy. The method was tested on QuickBird and IKONOS images and an overall classification accuracy of 92% and 86% was achieved for each of the images,

respectively. The method offers a practical, fast, and easy to use (within eCognition) framework for classifying VHR imagery of small urban areas.

Third, a combined object- and pixel-based image analysis framework is proposed to overcome the limitation of object-based (lack of general applicability and automation) and pixel-based (ignoring the spatial information of the image) approaches. The framework consists of three major steps: image segmentation, feature extraction, and pixel-based classification. For the feature extracting part, a novel approach is proposed based on the wavelet transforms. The approach is unsupervised and much faster than the current techniques because it has a local scope and works on the basis of an image's objects, not pixels. The framework was tested on WorldView-2, QuickBird, and IKONOS images of the same area acquired on different dates. Results show up to 17%, 10%, and 11% improvement of classification kappa coefficients compared to when only the original bands of the image are used for WorldView-2, QuickBird, and IKONOS, respectively.

Fourth, a novel object-based moving vehicle (and velocity) extraction method is developed using single WorldView-2 imagery. The method consists of three major steps: road extraction, moving vehicle change detection, and position and velocity estimation. Unlike recent studies in which vehicles are selected manually or semi-automatically using road ancillary data, the method automatically extract roads and moving vehicles using object-based image analysis frameworks. Results demonstrate a promising potential for automatic and accurate traffic monitoring using a single image of WorldView-2.

IN MEMORY OF MY FATHER

TO MY MOM

, and

TO MY WIFE

WITH LOVE AND ETERNAL APPRECIATION

ACKNOWLEDGEMENTS

I would like to take this opportunity to extend sincere thanks those people who made this work achievable. First, I would like to thank my supervisor, Dr. Yun Zhang for both his guidance and encouragement throughout my PhD program. He provided me many opportunities for teaching, attending conferences, and other academic activities. I would also like to thank my co-supervisor, Dr. Ming Zhong for all his valuable guidance and advice. I am very grateful to Dr. David J. Coleman and Dr. Susan E. Nicholas for their help in my major and minor examinations and for their review of my PhD research proposal and dissertation.

I would like to thank the City of Fredericton for providing QuickBird, IKONOS, and Spot Height data used in this research. Sincere thanks go to DigitalGlobe for providing WorldView-2 images used in this research.

Thanks are given to staff at the Department of Geodesy and Geomatics Engineering, especially David Fraser for all his non-stop technical support throughout my PhD study. I would also like to thank the university writing center, especially Dr. Richard Spacek for his valuable comments on my papers and dissertation. I also thank fellow graduate students at the Department of Geodesy and Geomatics Engineering.

I am truly grateful to my mother, and my late father, for each and every thing they did for me and for their care at every step of my life. I am also grateful to my father in law and my mother in law for their support and encouragement from thousands of miles away.

Finally, I would like to thank my beautiful wife, Farzaneh Shayanfar, for her unconditional love, support, patience, and understanding. Your spiritual support is the source of my energy ever!

Table of Contents

List of Tables

List of Figures

List of Symbols, Nomenclature or Abbreviations

VHR - Very High Resolution

FbSP - Fuzzy-based Segmentation Parameter

DSM – Digital Surface Model

DEM – Digital Elevation Model

DTM – Digital Terrain Model

PCA – Principal Component Analysis

MS – Multispectral

Pan – Panchromatic

WV-2 – WorldView-2

QB – QuickBird

IK – IKONOS

GIS – Geographic Information Systems

OBIA – Object-based Image Analysis

GEOBIA – Geographic Object-based Image Analysis

OA – Overall Accuracy

KC – Kappa Coefficient

PA – Procedure's Accuracy

UA – User's Accuracy

GLCM – Grey Level Co-occurrence Matrix

HOM – Homogeneity

ASM – Angular Second Moment

CON – Contrast

ENT – Entropy

IDM – Inverse Difference Moment

VAR – Variance

STD – Standard Deviation

DIS – Dissimilarity

MEN - Mean

NDVI – Normalized Difference Vegetation Index

nDSM – Normalized Digital Surface Model

SH – Spot Height

NDWI – Normalized Difference Water Index

FNEA – Fractal Net Evolution Approach

RPC – Rational Polynomial Coefficients

TP – True Positive

TN – True Negative

FP – False Positive

FN – False Negative

Chapter 1 : INTRODUCTION

This PhD dissertation presents research on the analysis of very high spatial resolution (VHR) optical satellite imagery for urban land cover classification and moving vehicle (and its velocity) extraction. It is an article-based dissertation presented through the following papers.

Paper 1 (Peer reviewed):
Salehi, B., Y. Zhang, M. Zhong, and V. Dey , 2012. A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery. *International Journal of GeoInformatics,*8(3):35-51.

Paper 2 (Peer reviewed):
Salehi, B., Y. Zhang, and M. Zhong, 2012. Object-based land covers classification of urban areas using VHR imagery and Height Points ancillary data, *Remote Sensing* , 4(8):2256-2276.

Paper 3 (Peer reviewed):
Salehi, B., Y. Zhang, and M. Zhong, 2012. A combined pixel-and object-based image analysis framework for urban land cover classification of VHR imagery, *Photogrammetric Engineering & Remote Sensing* (under review).

Paper 4 (Peer reviewed):
Salehi, B., Y. Zhang, and M. Zhong, 2012. Automatic moving vehicle information extraction from single-pass WorldView-2 imagery, *IEEE Journal of Selected Topic in Earth Observation and Remote Sensing,* 5(1): 135-145.

1.1 Dissertation Structure

This research is an article-based dissertation including six chapters. Chapter 1 provides the introduction of the research. The next four chapters (chapter 2 to chapter 5) present four peer reviewed journal papers listed above, which are either published or under review. In each of the four papers, the first author conducted the primary research, while the second author provided advice on the structure and the remaining authors provided input and assistance. Chapter 6 provides the summary and conclusion of this research. **Figure 1.1** illustrates the organization of this dissertation.

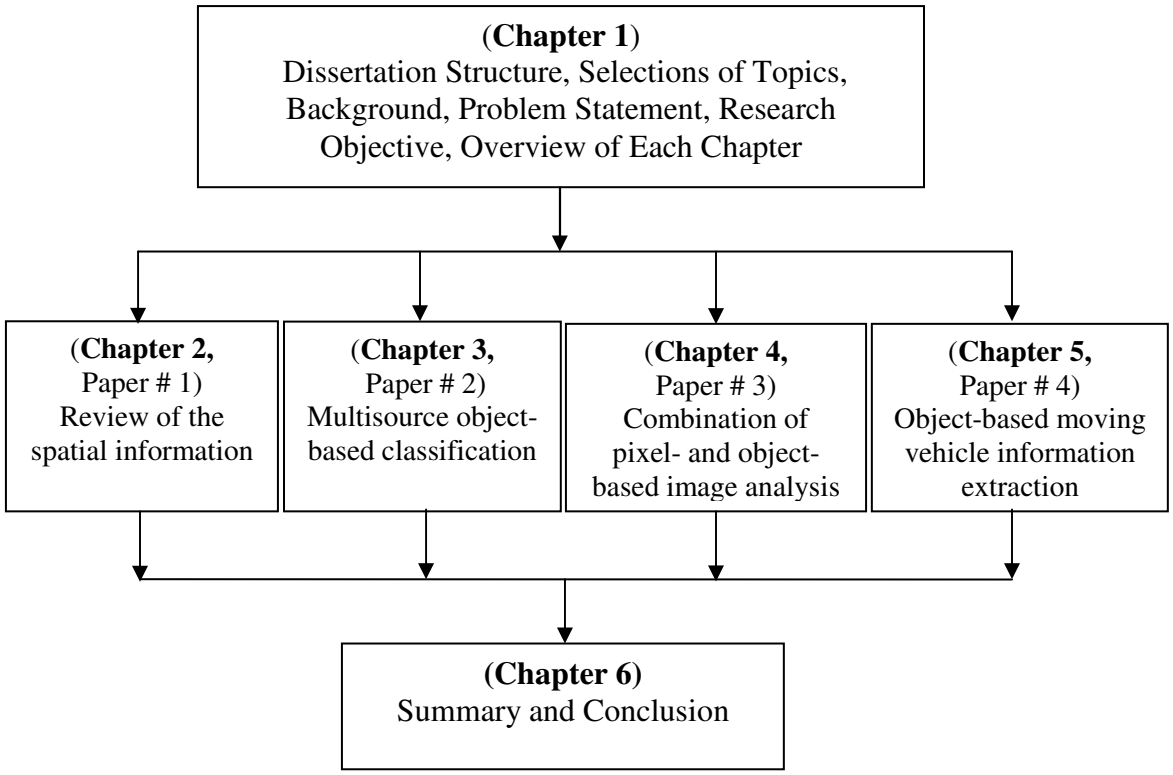


Figure 1.1 Structure of the dissertation

1.2 Selection of Research Topics

This PhD research is a part of an NSERC (National Sciences and Engineering Research Council) Strategic Project entitled "Development of a High Fidelity Urban Information System for Improved Land Use and Transportation Planning." The project is a collaborative research between the Departments of Geodesy and Geomatics Engineering (Remote Sensing group) and Civil Engineering (Transportation group) at the University of New Brunswick. The Remote Sensing Group is responsible for developing urban land cover classification/mapping techniques. The land use classification component is the responsibility of the Transportation Group in Civil Engineering.

Therefore, the primary objective of this PhD research is to investigate sophisticated algorithms and methods for urban land cover classification using very high spatial resolution (VHR) optical imagery. During the course of this research we learned that the VHR data used (particularly WorldView-2 imagery, which became available in the market in 2010) has potential for mapping and classification of smaller objects than land cover such as vehicles. In addition, vehicle extraction is of great interest to our partner research group (Transportation Group). Therefore, we further extended the topics of this research to moving vehicle (and its velocity) extraction in urban areas using VHR imagery. The next three chapters (chapters 2-4) concern urban land cover classification, while the fifth chapter is about the extraction of moving vehicles and velocity estimation.

1.3 Background

Since the launch of the first very high spatial resolution (VHR) optical satellite in 1999 (i.e., IKONOS), the analysis of images acquired by such satellites has received attention within the remote sensing community. VHR imagery normally possesses the spatial resolution of better than 1 meter for the panchromatic band and up to four meters for the multispectral bands. As Table 1.1 shows, over 10 VHR optical satellites are currently orbiting the Earth. This number is expected to nearly double over 2010-2015 period [Euroconsult, 2011].

Table 1.1 Very High Resolutions (VHR) Optical Satellite[1]

1. Satellite	*Country*	Launch	Pan Res.(m)	MS Res. (m)	# of MS Bands	Swath (Km)
GeoEye-1	*USA*	06/09/08	0.41	1.64	4	15
WorldView-1	*USA*	18/09/07	0.5	N/A	0	16
WorldView-2	*USA*	08/11/09	0.46	1.84	8	16
QuickBird	*USA*	18/10/01	0.6	2.5	4	16
EROS B1	*Israel*	25/04/06	0.7	N/A	0	7
KOMSAT-3	*Korea*	17/05/12	0.7	2.8	4	15
Pleiades-1	*France*	16/12/11	0.5	2.0	4	20
IRS Cartosat-2	*India*	10/01/07	0.8	N/A	0	10
IKONOS	*USA*	24/09/99	1.0	4.0	4	11
Resurs DK-1	*Russia*	15/06/06	1.0	3.0	3	28
KOMPSAT-2	*Korea*	28/07/06	1.0	4.0	4	15
OrbView-3	*USA*	26/06/03	1.0	4.0	4	8

Pan: Panchromatic; MS: Multispectral; Res.: Resolution

Over the past decades, the majority of remote sensing work has been focused on natural environment [Weng and Quattorchi, 2007]. Employing remote sensing imagery for urban applications has gained interest within the remote sensing community in recent years due to the greater availability of VHR satellite imagery. On the one hand, the sub-meter

[1] Modified from Stoney, W.E., Mitretek systems, 2008-2-12, http://www.asprs.org/Satellite-Information/

spatial resolution of VHR images makes them a suitable choice for detailed mapping of urban areas and vehicle extraction. On the other hand, the limited spectral resolution and, more importantly, the high spatial correlation between the adjacent pixels raise serious challenges for the analysis of such imagery if the conventional spectral- and pixel-based image analysis approaches are employed [Mohapatra and Wu, 2010; Myint *et al.,* 2011]. The primary objective of this research is to explore and investigate more sophisticated methods for **urban land cover classification** and **moving vehicle extraction** using VHR imagery.

1.3.1 Urban Land Cover Classification

There are two terms that are frequently used in the field of remotely sensed image classification, namely land cover and land use. Land cover refers to the type of material present on the landscape (e.g., water, crops, and buildings) while land use refers to what people do on the land surface (e.g., agricultural, commercial, and residential) (Jensen, 2005). Land cover is the basis of land use applications. In fact, if the land cover is accurately mapped and classified, the land use, which is more application-dependent, can be more accurately classified. For urban land use classification, additional information such as the zoning and spatial dependency (e.g., census information) is required. The concern of this research is land cover classification.

One of the most widely used applications of VHR imagery is mapping and classification of land covers in urban areas. However, because of the complexity of urban landscapes, the failure to consider spatial characteristics of the image in traditional pixel-based classifiers, and inconsistencies between scale of observation (i.e., pixel size) and the

spatial characteristics of targets (e.g., buildings), conventional spectral-based classification approaches are ineffective [Weng and Quattrochi, 2007]. Therefore, sophisticated techniques are required to incorporate not only the spectral information of the image, but also the spatial information (e.g., texture, shape, and size) into the classification process.

1.3.2 Moving Vehicle Information Extraction

Vehicle monitoring is one of the important issues for modeling and planning of traffic and for transportation management in both terrestrial and marine areas. Traditionally, traffic monitoring has been conducted using ground-based equipment such as inductive loops, radar systems, and video cameras [Munno *et al.*, 1993; Casstellano, 1999; Nag and Barnes, 2003]. These systems, however, have a very narrow coverage area and are not able to observe global traffic situations. The use of remote sensing data for traffic monitoring has emerged as an alternative in recent years. A few studies have used LiDAR data for monitoring moving vehicles [e.g., Toth and Grenjer-Brzeinska, 2006; Yao *et al.,* 2011). However, LiDAR data are expensive and do not have broad coverage areas. An alternative is to use VHR imagery.

One of the interesting applications of VHR imagery, which has gained attention in the remote sensing community in recent years, is to extract the moving vehicles together with their velocity information in urban areas [Xiong and Zhang, 2008; Liu *et al.,* 2010; Leitloff *et al.,* 2010]. With about half meter spatial resolution of VHR imagery (e.g., QuickBird and WorldView-2), small targets such as vehicles can be observed and detected. Further, because of the sensor configuration of VHR satellites such as

QuickBird and WorldView-2, a moving target and its velocity can be extracted. Most VHR satellites have a panchromatic and a multispectral sensor onboard (WorldView-2 has a panchromatic and two multispectral sensors). These two sensors (three in WorldView-2) acquire images with a small time lag (approximately 0.15 seconds for QuickBird). Consequently, if the target being observed by the satellite is in motion, the panchromatic and multispectral sensors record two different positions of the target. By comparing these two positions the moving vehicle and its velocity, including speed and direction, can be extracted.

1.4 Problem Statement

1.4.1 Urban Land Cover Classification of VHR Imagery

In a typical urban landscape there exist different impervious and pervious land cover classes. Impervious land covers include buildings (large buildings and small houses), roads (city streets and highways), and parking and paved areas. Pervious land covers are mainly vegetation (grass and trees), water, and soil. The classification of pervious surfaces using the spectral information of VHR imagery often results in high classification accuracies, i.e., more than 90%, due to the high spectral difference between such classes [Salehi *et al.,* 2012]. However, the problem arises when the objective is to extract impervious land cover classes. This is because of the within-class spectral variation and between-classes spectral similarity of such classes, particularly in VHR imagery.

Many researchers have investigated methods to tackle this problem. These methods often incorporate not only the spectral information of the image but also additional spatial information.

1.4.1.1 Incorporation of Spatial Information into Classification

Spatial information beyond the spectral information (brightness values) of individual pixels is required for classification of impervious surfaces of urban areas using VHR imagery. The spatial information is extracted either from ancillary data such as existing GIS vector layers and LiDAR data, or from the spectral and spatial correlation between neighboring pixels in the image. In the latter case, the information is categorized as three major types: textural, morphological (e.g., shape and size), and contextual information. Although the spatial information has been employed in individual publications, no comprehensive literature review has been conducted to date on the effectiveness of each individual type of information in urban land cover classification of VHR imagery. Therefore, a comprehensive study of the usefulness of such information in urban land cover classification of VHR imagery is required.

1.4.1.2 Multisource Object-Based Classification

As mentioned, additional information can be extracted from ancillary data. When ancillary data are combined with the image, the classification method is generally referred to as multisource classification [Watanchaturaporn *et al.,* 2008]. The ancillary data used are mostly GIS data layers and LiDAR height data. LiDAR data has been used in the literature in recent years for classification of VHR imagery of urban areas [Ye *et al.,* 2010; Wurm *et al.,* 2011; Hermosilla *et al.,* 2011]. However, LiDAR data are

expensive and may not be available for many urban environments. On the other hand, GIS data such as height points, building and road maps are often available from archives for most urban areas. Despite the availability of GIS data, they have rarely been utilized for detailed classification of urban areas using VHR imagery [Salehi *et al.,* 2012].

Among a few examples are Thomas *et al.,* [2003] and Zhou *et al.,* [2009]. However, these examples have been developed for specific applications and have not considered the general problems in multisource classification. While geometric misregistration between images and GIS vector layers poses a great challenge for integration, previous studies have not put much effort toward overcoming this issue. Object-based image analysis has proved to be more effective for multisource classification. However, the general applicability and transferability of the object-based classification framework is another issue which has gained less attention in the current literature. Therefore, an object-based image analysis framework that combines the VHR image and vector data, and is applicable to different VHR images, for detailed urban land cover classification is required.

1.4.1.3 Combination of Pixel- and Object-Based Image Analysis for Classification

Multisource classification is effective for classifying complex urban environments because of the incorporation of additional information to VHR images in the classification process. However, the success of such classification depends on the availability of ancillary data. For cases in which multisource data is not available, a classification approach which exclusively uses the VHR imagery is desirable. Such a classification approach incorporates additional information extracted from the image into

classification. An effective way of incorporating additional information into classification is using object-based image analysis. The first step in object-based image classification is segmentation, which is the process of partitioning the image into a set of discrete, non-overlapping regions on the basis of internal homogeneity criteria [Devereux *et al.,* 2004]. In the second step, classification is performed using the objects' properties such as spectral, textural, morphological, and contextual measures.

In recent years, object-based image analysis has proved to be very effective in urban land cover classification of VHR imagery and many publications have benefited from that [Blaschke, 2010]. However, a major problem associated with the object-based classification approach is its subjectivity to area and application (in most cases, it first need to be developed using a small part of the image and then applied to the entire image). In fact, object-based approaches are not automatic. On the other hand, traditional pixel-based classification approaches are more generic but ignores the spectral and spatial properties of pixels surrounding the target pixel in the classification process. Therefore, a combined object- and pixel-based method which takes the advantage of each approach is desirable.

By segmenting the image, a great number of object properties such as spectral (e.g., mean brightness value), textural (e.g., grey level co-occurrence matrix), and morphological (e.g., shape and size) features can be derived and utilized in a subsequent pixel-based classification. A necessary step in such a combination is to select/extract the best features from the huge number of derived features. This is because utilizing such an amount of data is computation-intensive, and more importantly, will lead to inaccurate classification

due to the curse of dimensionality [Bellman, 1961]. Therefore, an automatic and fast feature extraction algorithm is necessary for the combination of object-based and pixel-based image analysis.

1.4.2 Moving Vehicle Information Extraction

A few studies have used QuickBird imagery for extracting moving vehicle information [Xiong and Zhang, 2008; Pesaresi *et al.,* 2008; Liu *et al.,* 2010; Leitloff *et al.,* 2010]. The methods , however, required additional data such as road maps for identifying vehicles from the VHR image [Lie *et al.,* 2010; Leitloff *et al.,* 2010]. In some cases, the vehicles are selected manually followed by extracting velocity and position [Xiong and Zhang, 2008; Pesaresi *et al.,* 2008].

The new sensor configuration of WorldView-2 in addition to its high spectral resolution (eight spectral bands) has provided additional opportunities for moving vehicle extraction using VHR imagery. No previous study has benefited from these features of WorldView-2 imagery for vehicle extraction, given the very recent availability of Worldview-2 imagery in the market (since 2010). Therefore, it is desirable to develop a method exploiting the special characteristics of WorldView-2 imagery for moving vehicle extraction.

1.5 Research Objectives

The objectives of this research are fourfold, in order to solve the limitation identified in the above areas related to urban land cover classification and moving vehicle extraction from VHR imagery. Therefore, the four main objectives are as follows:

a) Provide a comprehensive literature review for evaluating the usefulness of textural, contextual, and morphological information of VHR imagery, and ancillary data (e.g., vector and LiDAR data) for urban land cover classification (both pixel-based and object-based).

b) Develop a multisource object-based classification framework using the combination of VHR imagery and vector ancillary data such as Spot Height data.

c) Develop a combined object- and pixel-based image analysis framework exploiting the advantages of both approaches for urban land cover classification of VHR imagery.

d) Develop an object-based moving vehicle extraction and velocity estimation method by exploiting the characteristics of a single image of the recently launched WorldView-2 satellite.

1.5.1 Review of the Spatial Information

To this author's knowledge, there has not been a comprehensive study on using spatial information, extracted either from the image or from ancillary data, in urban land cover classification of VHR imagery. The objective of **Chapter 2** is to provide such a review. In particular, the objective is to evaluate the usefulness of each type of spectral, textural, morphological, contextual, and ancillary information in pixel-based and object-based classification of VHR imagery over urban areas. The contribution of each type of information in increasing the classification accuracy is quantitatively evaluated by reporting accuracies achieved in some recent publications. Another objective is to identify which type of spatial information has not been extensively utilized in the literature and need further investigation. This literature review is the basis of the

subsequent chapters in this dissertation; it also provides guidelines for future research in urban land cover classification of VHR imagery.

1.5.2 Multisource Object-Based Classification

From the literature review [Salehi *et al.,* 2012], it has been identified that very few studies have been conducted on multisource urban land cover classification of VHR imagery. In addition, object-based approach has proved to be more effective for multisource classification. The objective of **Chapter 3** is to develop a multisource object-based classification framework using height point data and VHR imagery of urban areas. In particular, the research assesses the general applicability and transferability of the developed framework to different areas and VHR images. It also evaluates the effect of possible misregistration between vector data and VHR images on object-based classification.

1.5.3 Combination of Object- and Pixel-Based Image Analysis

The general objective of **Chapter 4** is to develop a methodology exploiting the advantages of object-based image analysis (the use of additional spectral and spatial features of objects) and pixel-based classification (being automatic and independent of the area, application, and imagery) approaches. Specifically, the method focuses on developing a novel feature extraction method to conquer the problem of curse of dimensionality [Bellman, 1961] resulting from the huge number of object features derived from segmenting the VHR image. For this, a novel wavelet-based feature extraction method is developed. The proposed feature extraction method works on the basis of objects, as oppose to current feature extraction techniques which work on the

basis of pixels, thus the computation is dramatically reduced compared to the conventional feature extraction methods such as principal component analysis. Another objective is to test whether the additional bands of WorldView-2, compared to the conventional bands available in QuickBird and IKONOS, have significant impact on increasing the classification accuracy of impervious land cover types. Another objective is to test which set of objects features (i.e., spectral, textural, and morphological features) contributes more to increasing the classification accuracy. Finally, the last objective is to evaluate the impact of object size (level of segmentation) on the classification performance.

1.5.4 Automatic Moving Vehicle Information Extraction

Vehicle extraction and velocity estimation using a single VHR image has been considered in some recent studies (see section 1.4.2). However, the 0.5 m spatial resolution, the four new multispectral bands in addition to the four standards multispectral bands, and the new sensor configuration (i.e., four multispectral bands on either side of the panchromatic sensor) in the newly launched WorldView-2 have increased the opportunity for automatic moving vehicle information extraction. The objective of **Chapter 5** is, thus, investigating an automatic method for extracting the moving vehicles and their information including position, speed, and direction using a single WorldView-2 image.

1.6 Data and Metrics

The data and metrics used to evaluate the methods developed in each chapter are summarized in **Table 1.2**.

Table 1.2 Data and metrics used for evaluation of the proposed methods

Source of data	Data	Metric description	Chapter
QuickBird, IKONOS and Height Point were provided by City of Fredericton	(1) IKONOS and QuickBird images of Fredericton, New Brunswick, acquired in 19 June and 31 August 2002, respectively. Each image contains 1 panchromatic band and four multispectral bands. (2) a layer of height points called Spot Height layer collected from 2002 stereo aerial photography of scale of 1:10000 from City of Fredericton (3)Two manually digitized maps for the two images. The maps used as references for accuracy assessment	*Confusion matrix* (Congalton and Green 2009) and its all elements including *overall accuracy*, *kappa coefficient*, *user's* and *producer's accuracies*. The *Z statistics* of each matrix	Chapter 3 (Salehi *et al.,* 2012)
The WorldView-2 image was provided to the author through "ERDAS IMAGINE-DigitalGlobe 2012 Geospatial Challenge" contest	(1) IKONOS, QuickBird and WorldView-2 images of Fredericton acquired in June 19, 2002, August 31, 2002, and July 20, 2011, respectively. WorldView -2 contains a panchromatic band with spatial resolution of 0.5 m and eight multispectral bands of 1.8 m resolution. (2) Three manually digitized maps for the three images. The maps used as references for accuracy assessment	*Confusion matrix* (Congalton and Green 2009) and its all elements including *overall accuracy*, *kappa coefficient*, *user's* and *producer's accuracies*.	Chapter 4 (Salehi *et al.,* 2012)
The WorldView-2 image was provided to the author through "DigitalGlobe 8-band Research Challenge" contest	WorldView-2 image of Moncton, New Brunswick, acquired in October 5, 2010. The image contains a panchromatic bands and two sets of multispectral bands (each contains four bands). The image covers a low speed traffic zone and a high speed traffic zone of Moncton.	*Correctness, Completeness, and Quality measures* (Agouris *et al.,* 2004)	Chapter 5 (Salehi *et al.,* 2012)

1.7 Overview of Each Chapter

Chapter 1 presents the introduction. It comprises the structure of dissertation, research topic selection, research background, problem statement, objectives of the research, and an overview of each remaining chapter. `

Chapters 2 to 5 contain the four journal papers representing the main contributions of this PhD dissertation.

- **Chapter 2** presents a comprehensive review of recent literature in the field of urban land cover classification using VHR imagery. It provides research direction for chapter 3-4.
- **Chapter 3** demonstrates a new hierarchical rule-based and object-based classification framework using the combination of VHR imagery of urban areas.
- **Chapter 4** introduces a combined object- and pixel-based image analysis framework for urban land cover classification of VHR imagery. The framework includes three major parts, multiresolution segmentation, wavelet-based feature extraction, and pixel-based classification.
- **Chapter 5** presents a novel object-based method for automatic extraction of moving vehicles from a single set of WorldView-2 imagery. The method contains three major steps, object-based road extraction, moving vehicle extraction, and velocity estimation.
- **Chapter 6** presents the conclusions and recommendations. It summarizes the achievements of this dissertation followed by the suggestions for future work.

REFERENCES

Agouris, P., Doucette, P. and Stefandis, A. [2004]. Automation and digital photogrammetric workstations, In: McGlone, J.C., Mikhail, E.M., Bethel, J., Mullen, R. , Manual of Photogrammetry, fifth ed. American Society for Photogrammetry and Remote Sensing, Bethesda, MA, pp. 949-981.

Bellman, R. [1961]. *Adaptive Control Processes: A Guided Tour*, Princeton University Press, Princeton, New Jersey, 255 p.

Blaschke, T. [2010]. Object-based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 65, No. 1, pp. 2-16.

Castellano, G., Boyce, J. and Sandler, M. [1999]. CDWT optical flow applied to moving target detection, *IEE Colloquium on Motion Analysis and Tracking*, pp.17/1-17/6.

Congalton, R.G., and Green, K. [2009]. *Assessing the accuracy of remotely sensed data: principles and practices,* Second edition. Taylor & Francis Group, Boca Raton, Florida, 183 p.

Devereux, B.J., Amable, G.S. and Posada, C.C. [2004]. An efficient image segmentation algorithm for landscape analysis. *International Journal of Applied Earth Observation and Geoinformation*, 6(1), 47-61.

Euroconsult© [2011). "Earth Observation Industry Remains Buoyant Despite Challenging Economic Environment", Euroconsult© News available online at ttp://www.euroconsult-ec.com/news/press-release-33-1/52.html.

Hermosilla, T., Ruiz, L. A.Recio, J. A. and Estornell, J. [2011]. Evaluation of Automatic Building Detection Approaches Combining High Resolution Images and LiDAR Data. *Remote Sensing,* Vol. 3*,* No. 6, pp. 1188-1210.

Jenson, R. J. [2005]. *Introductory Digital Image Processing: A Remote Sensing Perspective.* Pearson Prentice Hall, NJ, USA.

Leitloff, J. ,Hinz, S. and Stilla,U. [2010]. Vehicle detection in very high resolution satellite images of city areas, *IEEE Trans. Geosci. Remote Sens.* vol.48, no.7, pp.2795-2806.

Liu, W., Yamazaki, F. and Vu, T.T. [2010]. Automated Vehicle Extraction and Speed Determination From QuickBird Satellite Images, *IEEE J. Selected Topics in Applied Earth Observations and Remote Sens.*, vol.4, no.1, pp.75-82.

Mohapatra, R.P. and Wu, C. [2010]. High resolution impervious surface estimation: an integration of Ikonos and Landsat-7 ETM+ imagery. *Photogrammetric Engineering and Remote Sensing*, 76(12), 1329-1341.

Munno, C. J., Turk, H., Wayman, J. L. , Libert, J. M. and Tsao, T. J. [1993]. Automatic video image moving target detection for wide area surveillance," *In Proc. Int. Con. Security Technology,* pp.47-57.

Myint, S.W., Gober, P., Brazel, A., Clark, S. G. and Weng, Q. [2011]. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment,* 115(5), 1145-1161.

Nag, S. and Barnes, M. [2003]. A moving target detection filter for an ultra-wideband radar, *In Proc. IEEE Radar Conference*, pp. 147- 153, 2003.

Pesaresi, M., Gutjahr, K.H. and Pagot, E. [2008]. Estimating the velocity and direction of moving targets using a single optical VHR satellite sensor image, *Int. J. of Remote Sens.*, vol. 29, no.4, pp. 1221-1228.

Salehi, B., Zhang,Y., Zhong, M. and Dey, V. [2012]. A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery. *International Journal of GeoInformatics,* 8(3):35-51.

Thomas, N., Hendrix, C. and Congalton, R.G. [2003]. A comparison of urban mapping methods using high-resolution digital imagery. *Photogrammetric Engineering and Remote Sensing*, Vol. 69, No. 9, pp. 963-972.

Toth, C. K. and Grejner-Brzezinska, D. [2006]. Extracting dynamic spatial data from airborne imaging sensors to support traffic flow estimation, *ISPRS Journal of Photogram.m Remote Sens*., vol. 61, no. 3/4, pp. 137–148.

Watanachaturaporn, P., Arora, M. K. and Varshney, P. K. [2008] Multisource Classification using Support Vector Machines: An Empirical Comparison with Decision Tree and Neural Network Classifiers. *Photogrammetric Engineering and Remote Sensing*, Vol. 74, No. 2, 239-246.

Weng, Q. and Quattrochi, D. A. [2007]. *Urban Remote Sensing,*CRC Press, Taylor& Francis Group, Boca Raton, 412 p.

Wurm, M.,Taubenböck, H., Schardt, M.,Esch, T., and Dech, S. [2011]. Object-based image information fusion using multisensor earth observation data over urban areas. *International Journal of Image and Data Fusion*, Vol. 2, No. 2, pp. 121 – 147.

Xiong, Z. and Zhang, Y. [2008]. An initial study on vehicle information extraction from single pass of satellite QuickBird imagery," *Photogramm. Eng. Remote Sens*., vol. 74, no. 11, pp. 1401–1411

Yao, W., Hinz, S. and Stilla,U. [2011]. Extraction and motion estimation of vehicles in single-pass airborne LiDAR data towards urban traffic analysis," *ISPRS J. of Photogramm. Remote Sens.,* vol. 66, no.3,pp. 260-271.

Yu, B. Liu, H, Wu, J., Hu, Y. and Zhang, L. [2010]. Automated derivation of urban building density information using airborne LiDAR data and object-based method. *Landscape and Urban Planning,* Vol. 98, No. 3-4, pp. 210–219.

Zhou, W., Huang, G., Troy, A. and Cadenasso, M.L. [2009]. Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: A comparison study. *Remote Sensing of Environment*, Vol. 113, No. 8, pp. 1769-1777.

Chapter 2 : A REVIEW OF THE EFFECTIVENESS OF SPATIAL INFORMATION USED IN URBAN LAND COVER CLASSIFICATION OF VHR IMAGERY[1]

Abstract

Land cover classification of very high resolution (VHR) imagery in urban areas is an extremely challenging task, because of the low intra-class (within-class) and high inter-class (between-classes) spectral similarities of impervious land cover types (such as buildings and traffic areas). Over the past decade, a significant amount of research has been conducted on the incorporation of spatial information along with spectral information of VHR imagery into urban land cover classification. The spatial information includes textural, morphological and contextual measures extracted from VHR imagery, as well as LiDAR- and photogrammetrically-derived DSM and existing GIS data layers. In this paper, a comprehensive review of recent literature was conducted to evaluate the effectiveness of such measures in land cover classification of urban areas using VHR imagery. For each measure, a comprehensive list of papers for both pixel-based and object-based classification is provided. In addition, the classification results of

[1] This paper has been published in the *International Journal of Geoinformatics:*
Salehi, B., Zhang,Y., Zhong,M., Dey, V., 2012. A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery. *International Journal of GeoInformatics,*8(3):35-51.

For consistency throughout the dissertation, the format and style of Figures' captions, Table's titles, citation of references in the text, and section numbering have been slightly changed (from the original format of the journal in which paper has been published or is under review) for chapters 2-5.

representative publications are reported for each measure and its advantages and limitations in both pixel-based and object-based approaches are discussed. It has been found that, in general, object-based classification performs better than pixel-based approaches, since it facilitates the use of spatial measures by segmenting the image. Moreover, utilizing spatial measures significantly improves the classification performance for impervious land cover types, while may have no effect or even lower the classification accuracy for classes of vegetation and water surfaces. Textural measures are more commonly utilized in pixel-based approaches, while morphological measures have better performance in object-based classification. The effect of contextual measures on classification is enhanced when these measures are used in conjunction with two other measures, particularly in object-based approaches. Although ancillary data shows a very high potential to address the problem of spectral-based classifiers in separating spectrally similar impervious land cover types, incorporating such data, particularly photogrammetrically-derived DSM, in classification is still in a very early stage and requires significant exploration and development.

2.1 Introduction

Land cover classification is one of the most important topics in remote sensing both for researchers and practitioners, because of its broad applications in almost all geo-related domains. Remotely sensed images are the major, and sometimes the only, input in land cover classification. The spatial resolution of the image is one of the most important factors that affect land cover classification performance [Chen *et al.,* 2004]. Previous

research has explored the impact of spatial resolution on classification of remotely sensed data [e.g., Price, 1997; Quattrochi and Goodchild, 1997].

Because of the sub-meter ground resolution, VHR images unveil a very high potential for more detailed and accurate mapping of the urban environment [Pacifici *et al.,* 2009]. New VHR digital aerial cameras provide an excellent data source for the mapping and classification of urban areas, but their images are expensive and not easy to collect in a short period of time. With the advent of very high spatial resolution (≤1m) satellite sensors since 1999, such as IKONOS, QuickBird, OrbView, WorldView-1, GeoEye-1 and WorldView-2, urban land cover classification has rapidly gained interest within the remote sensing community. However, the increased spatial resolution of VHR imagery does not automatically yield improved accuracy of urban land cover classification, if classifiers just employ the spectral information of the image (spectral-based classifiers). This is mainly due to the high spectral variation within the same land cover (intra-class spectral variation; e.g., buildings with different roof types) and the spectral confusion between different land covers (inter-class spectral confusion) [Lu *et al.,* 2010; Xu and Li, 2010; Huang *et al.,* 2011].

To compensate for the limitations of spectral-based classifiers, many researchers have attempted to develop techniques to incorporate spatial information extracted from VHR imagery and/or from ancillary data into classification. For the sake of convenience, we categorize the spatial information into four types: textural, contextual and morphological measures (extracted from the VHR image), and ancillary data such as digital

elevation/surface model (DEM/DSM) derived from LiDAR or stereo photographs and existing GIS data layers (e.g., road network and buildings' footprint).

This paper aims to review the most recent research on urban land cover classification using VHR satellite imagery. Specifically, the objective is to evaluate the effectiveness of incorporating the four aforementioned types of spatial measures in both pixel-based and object-based classification approaches. Comprehensive lists of major publications (including peer-reviewed journal papers) in which these four types of spatial measures have been utilized in both pixel-based and object-based classification approaches are reported. In addition, the effect of each type of measures on increasing the classification accuracy is quantitatively evaluated by reporting the accuracies achieved in some recent literature.

To date, we have not found a comprehensive review of different spectral and spatial measures used in classification of VHR imagery, particularly over urban areas. Lu and Weng [2007] conducted a survey of image classification methods and techniques for improving classification performance. Their survey includes a brief description of general process of image classifications, with the citation of a large amount of previously published literature. However, it did not focus on the classification of VHR imagery in urban environment. Liu *et al.,* [2006] briefly reviewed classification patterns of remotely sensed imagery based on object-oriented approaches. Their study, however, is limited to describing basic steps of object-oriented image analysis along with reviewing a few selected publications. Gamba *et al.,* [2005] presented a bibliographic review of the state-of-the-art of urban remote sensing using multiple data sets. They briefly reviewed the

data fusion issues in urban areas without taking into consideration the capabilities of VHR imagery in urban analysis.

2.2 Spatial Information Used in Classification

The aforementioned spatial measures are utilized in either pixel-based or object-based classification in order to help the classifier distinguish different land cover classes. Since the nature of such measures and their effects on classification are different in pixel-based and object-based image analysis, it is useful to give a brief review of these two generic types of classification before proceeding to the review of the measures.

2.2.1 Pixel-based vs. Object-based Classification

In general, image classification approaches can be grouped into different categories such as supervised or unsupervised, parametric or non-parametric, hard or soft (fuzzy) [Lu and Weng 2007]. In each category the basic processing unit could be pixel or object; accordingly approaches are described as pixel-based (or per-pixel) and object-based.

Although pixel-based approaches are still widely used for mapping particular urban impervious land cover types such as large commercial parcels, the distribution and shape of such cover types in heterogeneous areas may be more accurately mapped by object-based classification approaches [Hester *et al.,* 2008]. A serious problem associated with pixel-based classifiers is the so-called “salt and pepper” effect or “structural clutter”[Van de Voorde *et al.,* 2007], which produces a noisy classification result due to the spectral heterogeneity of classes in an urban environment and the lack of topological information used to classify pixels [Sims and Mesev, 2007]. This effect is more significant in the

VHR image classification especially over complex urban environments, because of the high intra-class spectral heterogeneity of some surface types. To reduce this negative effect of pixel-based classification approaches, some studies have proposed post-classification techniques [e.g., Van de Vorde *et al.,* 2007; Hester *et al.,* 2008]. For example, Hester *et al.* (2008) applied a 3x3 majority filter to eliminate the cluttered pixels resulted from the pixel-based classification. Van de Voorde *et al.* (2007) applied rules based on the contextual information encompass the pixel to eliminate the clutter (e.g. pixels that are completely surrounded by water are assigned to water).These techniques, however, may remove small land cover types such as single-family houses and single trees.

The object-based classification approach, on the other hand, decreases variance within the same land cover type by averaging the pixels within the object, which prevents the "salt and pepper" effect of pixel-based classification approaches [Chen *et al.,* 2009a]. Starting from around the year 2000, studies of object-based images analysis have sharply increased [Blaschke, 2010]. A comprehensive list of recent literature concerning object-based urban land cover classification of VHR imagery is provided in Tables 2.1, 2.4, 2.6, and 2.8. Other sources of object-based image analysis research include three online archives of conference proceedings (Object-based Image Analysis [OBIA 2006], GEographic Object-based Image Analysis [GEOBIA 2008 and GEOBIA 2010]), a book published on object-based image analysis by Blaschke *et al.,* [2008] and a literature review paper by Blaschke [2010]. It is noteworthy that much of the work referring object-based image analysis originated around the "eCognition" software [Benz *et al.,* 2004].

The next section reviews the spatial measures utilized in classification of land covers in a typical urban environment. The measures incorporated into both pixel-based and object-based classification are discussed and the advantages and limitations of each type of measure are evaluated.

2.2.2 Spectral-based Classifiers

Spectral-based classifiers have promising performance when applied to the medium and high spatial resolution images with several spectral bands for mapping relatively large homogeneous areas such as vegetation, forest, water and soil [Lu and Weng, 2007; McMahon, 2007; Xu and Li 2010]. However, because of similarities in the spectral response of land cover types in an urban scene, along with the low spectral resolution of VHR imagery (the limited number of spectral bands and the wide wavelength range covered by them), the classification of such imagery is a challenging task [Zhang and Couloigner, 2006].

Despite the fine spatial resolution of VHR imagery, its spectral resolution is limited to four multispectral bands (except the newly launched WorldView-2, which has eight multispectral bands) and a panchromatic band. Moreover, most of the bands that are suitable for separating urban land cover types lie outside or near the boundaries of the wavelength range of the multispectral bands of VHR imagery [Herold *et al.,* 2003b] (e.g., at wavelength around 580 nm which lies in the boundaries of the Green band or at wavelengths around 740nm which lies outside of Red and NIR bands of VHR images). Ben-Dor [2001], Herold *et al.,* [2004] and Warner and Nerry [2009] concluded that the shortwave and thermal infrared spectral regions are important for urban applications.

These bands, however, are not present in the VHR imagery. Thomas *et al.,* [2003] stated that the lack of mid-infrared bands in VHR images hinders the ability of traditional spectral-based classifiers to accurately distinguish detailed land cover types. Herold *et al.,* [2003b] also concluded that some land cover classes such as asphalt road, tar roof and parking lot have a very similar and constant low reflectance over the whole spectral range such that even the AVIRIS hyperspectral sensor has limitation in mapping these classes.

2.2.3 Spatial Measures Extracted from the Image

Spectral-based classifiers detect land cover classes exclusively according to spectral information while the large amount of valuable image spatial information is neglected. Moreover, in an urban landscape, impervious classes are spectrally too similar to be distinguished using only spectral information of the image. Hence, for the mapping of such classes, it is necessary to incorporate spatial information together with spectral information in the classification process. Two distinct types of method which utilize spatial information from an image are region-based and window-based methods [Gong *et al.,* 1992]. The region-based method is usually used in object-based, whereas the window-based method is used in pixel-based approaches.

In the following sub-sections, the performance of each type of spatial measures, i.e., textural, contextual and morphological measures of the image and spatial measures of ancillary data in both pixel-based and object-based approaches is reviewed. According to the results achieved in individual publications, the strengths and limitations of each group are discussed. The increase of classification accuracy resulting from the incorporation of spatial information into classification over conventional spectral-based classifiers is

reported. The classification accuracies reported through this review paper are all based on Error or Confusion matrix [Congalton, 1991; Richards and Jia, 2006] unless otherwise specified. For most literature, mentioned in tables, the Overall Accuracy (OA) and/or Kappa coefficient (KA) are used. In some cases, where available, the Producer's Accuracy (PA) of the classification of land covers such as buildings and roads is reported as well.

2.2.3.1 Textural Measures

Many researchers have attempted to employ texture measures as additional spatial information in the urban land cover classification of VHR images to overcome the lack of spectral information [Carleer and Wolff, 2006; Myint, 2007] in both pixel-based and object-based classification approaches (Table 2.1). The result of our literature review indicates that texture features extracted from gray level statistics, especially those from co-occurrence gray level matrix (GLCM), are the most useful texture measures used in the classification of VHR images over urban areas, particularly when pixel-based approaches are utilized.

Table 2.1 List of papers which have utilized different GLCM textural measures in urban land cover classification of VHR imagery

Pixel-based classification	Object-based classification
Zhang (1999); Pesaresi (2000); Maillard (2003); Shackelford and Davis (2003a); Shackelford and Davis (2003b); Chen *et al.,* (2004); Myint *et al.,* (2004); Walter (2004); Mena and Malpica (2005); Myint and Lam (2005); Puissant *et al.,* (2005); Zhang and Couloigner (2006); Alonso *et al.,* (2007); Myint (2007); Agüera *et al.,* (2008); Aksoy *et al.,* (2009); Pacifici *et al.,* (2009); Luo and Mountrakis (2010); Lu *et al.,* (2010); Ouma *et al.,* (2010); Tassetti *et al.,* (2010).	Herold *et al.,* (2003a); Thomas *et al.,* (2003); Walter (2004); Song *et al.,* (2005); Carleer and Wolf (2006); Su *et al.,* (2008); Chan *et al.,* (2009); Hermosilla *et al.,* (2011); Pu *et al.,* (2011); Salehi *et al.,* (2011a).

Several recent publications have benefited from the GLCM for the purpose of mapping urban areas using VHR imagery. For instance, Maillard (2003) concluded that GLCM gives superior results over the semivariogram and the Fourier-based texture extraction for the scenes where objects are distinguishable visually by their textures' characteristics. Buildings and traffic areas are well-textured classes in VHR images and they can be easily distinguished by visual interpretation. Consequently, the GLCM works better than the two others in this case.

Among the 14 GLCM texture measures, originally proposed by Haralick [1979], some of them are strongly correlated with each other [Cossu, 1988]. Thus, the choice of optimal texture measure is an important issue in GLCM texture extraction [Jensen, 2005]. Maillard [2003] reported that the most commonly used GLCM texture features in literature are, in decreasing order of popularity, the angular second moment (ASM), entropy (ENT), the inertia (initially contrast (CON)), the correlation (COR) and the inverse difference moment (IDM). Pacifici *et al.,* [2009] believe that energy (ENR) (which is the square root of ASM), CON, variance (VAR), COR, ENT and IDM are the most relevant measures used in literature. Based on the prototype performance approach and its application in urban areas [Pratt 2007], Puissant *et al.,* [2005] and Su *et al.,* [2008] concluded that four GLCM texture features used in mapping urban areas are homogeneity (HOM), ENT, dissimilarity (DIS) and the ASM.

We conducted a broad search on the major publications (mostly peer reviewed journals) in the area of urban land cover classification using VHR imagery to find which GLCM measures have mostly been utilized. The results are summarized in Table 2.2. Based on

the number of papers listed in Table 2.2 for each measure, a graph was plotted and is shown in Figure 2.1. ENT and ASM (or ENG) are the most frequently used measures. Nineteen and sixteen papers have utilized ENT and ASM respectively. HOM and CON ranked third (Figure 2.1). These two measures have been utilized in thirteen papers.

Table 2.2 List of papers which have utilized different GLCM textural measures in urban land cover classification of VHR imagery

GLCM	Paper
HOM	Zhang (1999); Herold *et al.,* (2003a); Puissant et al., (2005); Carleer and Wolf (2006); Agüera *et al.,* (2008); Su *et al.,* (2008); Pacifici *et al.,* (2009); Lu *et al.,* (2010); Luo and Mountrakis (2010); Ouma *et al.,* (2010); Tassetti *et al.,* (2010); Pu *et al.,* (2011); Salehi *et al.,* (2011a).
ASM (ENG)	Zhang (1999); Pesaresi (2000); Herold *et al.,* (2003a); Maillard (2003); Myint *et al.,* (2004); Puissant et al., (2005); Carleer and Wolf (2006); Myint (2007); Agüera *et al.,* (2008); Su *et al.,* (2008); Pacifici *et al.,* (2009); Luo and Mountrakis (2010); Lu *et al.,* (2010); Ouma *et al.,* (2010); Pu *et al.,* (2011); Salehi *et al.,* (2011a).
ENT	Zhang (1999); Pesaresi (2000); Herold *et al.,* (2003a); Maillard (2003); Shackelford and Davis (2003a); Shackelford and Davis (2003b); Myint *et al.,* (2004); Puissant et al., (2005); Carleer and Wolf (2006); Alonso *et al.,* (2007); Myint (2007); Agüera *et al.,* (2008); Su *et al.,* (2008); Chan *et al.,* (2009); Pacifici *et al.,* (2009); Ouma *et al.,* (2010); Tassetti *et al.,* (2010); Pu *et al.,* (2011); Salehi *et al.,* (2011a).
CON (Inertia)	Zhang (1999); Pesaresi (2000); Herold *et al.,* (2003a); Maillard (2003); Carleer and Wolf (2006); Myint (2007); Agüera *et al.,* (2008); Su *et al.,* (2008);Pacifici *et al.,* (2009); Luo and Mountrakis (2010); Ouma *et al.,* (2010); Pu *et al.,* (2011); Salehi *et al.,* (2011a).
COR	Maillard (2003); Alonso *et al.,* (2007); Myint (2007); Agüera *et al.,* (2008); Pacifici *et al.,* (2009); Ouma *et al.,* (2010); Pu *et al.,* (2011); Salehi *et al.,* (2011a).
IDM	Pesaresi (2000); Maillard (2003); Myint (2007).
VAR (STD*)	Herold *et al.,* (2003a); Shackelford and Davis (2003b); Agüera *et al.,* (2008); Chan *et al.,* (2009); Ouma *et al.,* (2010); Tassetti *et al.,* (2010); Pu *et al.,* (2011); Salehi *et al.,* (2011a).
DIS	Herold *et al.,* (2003a); Puissant et al., (2005); Carleer and Wolf (2006); Agüera *et al.,* (2008); Pacifici *et al.,* (2009); Lu *et al.,* (2010); Ouma *et al.,* (2010); Pu *et al.,* (2011).
MEN†	De Martino *et al.,* (2003); Agüera *et al.,* (2008); Chan *et al.,* (2009); ; Lu *et al.,* (2010); Ouma *et al.,* (2010); Tassetti *et al.,* (2010).

*STD: Standard Deviation ; †MEN:Mean

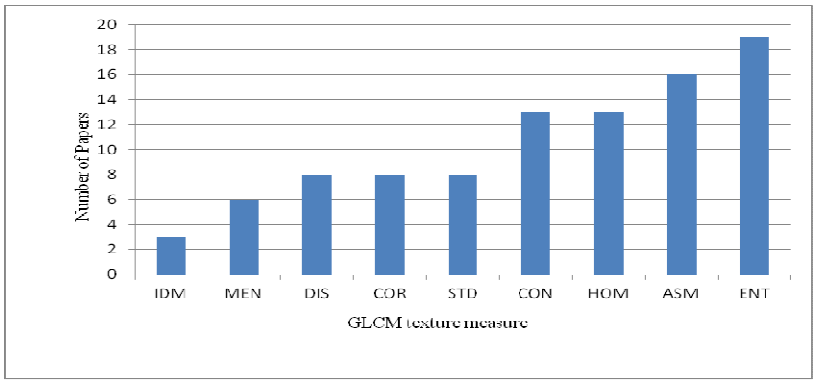


Figure 2.1 The most used GLCM texture measures for urban land cover classification of VHR imagery

Having determined the appropriate GLCM measure(s), four important factors including window size and shape, quantization level, inter-pixel distance and direction of spatial relationship, must be defined for each measure [Su *et al.,* 2008], since they influence the effectiveness of extracted measures in the classification process. The success of classification using texture features depends largely on the selected window size [Su *et al.,* 2008]. The window size and shape are automatically identified in object-based texture extraction inasmuch as the measure is calculated for each individual object resulting from the segmentation stage. In pixel-based approaches, however, the optimum of window size and three other factors must be determined. Window size is related to image resolution and content. Puissant *et al.,* [2005] believe that it would be interesting to choose different window sizes according to the size of features to be extracted. For the choice of direction, the literature proposes calculation of texture measures in four directions (0, 45, 90 and 135 degrees) and then taking the average of them [Haralick 1979; Anys *et al.,* 1994]. Several researchers suggest that it is a good idea to reduce the

quantization level (e.g., from 8 bits to 5 bits) of the input data so that the GLCM to be computed for each pixel does not become too large [Jensen, 2005].

Table 2.3 shows different texture measures with their corresponding texture factors used in some recent papers. Because of space limitations, we listed the results for the journal papers in which different VHR imagery, different urban land cover types, and different classification methods are utilized. From Table 2.1, it can be seen that the texture measures and their optimal factors differ from each other for different input data and different land cover types. Furthermore, the improvement in OA, as the result of incorporating texture into classification, ranges from 2% to 11% for papers listed in this table.

Table 2.3 Examples of using textural measures (with their corresponding optimal factors) for improving the urban land cover classification of VHR imagery. The textural measures are incorporated as additional band into the classification. Accuracy improvement (Imp) is over the case where only spectral band(s) of the image is utilized in the classification process.

Input data	Land cover types	Texture measures	Optimal Texture factors	Classification approach	Accuracy	Paper
VHR – Simulated image (1 pan and 3 MS bands at 1m Spatial res.)	Water ,shadow, tree, grass, road, built-up	HOM of pan band	W.S: :7x7 Dir.: average of (0,45,90, 135) I.P: 1 ; Q.L: NM	Pixel-based Discriminant Analysis	OA:92.2% Imp:4.4%	Puissant *et al.,* (2005)
VHR-QB (Pansharp bands)	Shrub, grassland, water, road, building, vacant land, shadow	ASM over objects from segmented image	Dir.: 45 Q.L:NM, I.P:NM	Object-based Maximum Likelihood	OA:83.7% Imp:2.1%	Su *et al.,* (2008)
		CON over image*	W.S:7x7 Q.L:NM, I.P:NM		OA:87.3% Imp:5.7%	
VHR-IK (Pansharp bands)	Road, building, tree, water, grass, shadow, bare soil	1st order statistic (ENT) of Pan band	W.S: 20x20 Dir., I.P, Q.L: Not Applicable	Pixel-based Hierarchical Fuzzy	Road-Buil.:73% Imp :1.5% Grass-Tree: 97.3% Imp :11%	Shackelford and Davis (2003b)
VHR-IK (Pan & MS)	Vegetation, soil, asphalt, metallic roof, shadow	ENT and COR of pan band	W.S:NM Dir. PC1 of 8 PCs for (0, 45, 90,135) Q.L:NM, I.P:NM	Pixel-based Maximum Likelihood	OA: 82.7% Imp: 3.4%	Alonso *et al.,* (2007)

*First, the texture measure was calculated for four MS bands resulting four textural bands, then Principal Component Analysis (PCA) was applied to four textural bands and the 1st PC was chosen as the final texture measure.
† Different combinations of W.S, Dir. and I.P for all five measures were used. For the reason of limited space it is not possible to present all of them in this table. The reader is referred to the cited paper for more information.
Acronyms definition: WV1=World View-1 imagery, Pan= Panchromatic, MS= Multispectral, W.S=Window Size, Dir=Direction, Q.L=Quantization Level, I.P=Inter Pixel distance, NM=Not Mentioned. Other acronyms are defined either in the text or in Table 2.1.

2.2.3.2 Contextual Measures

Whereas texture is the spatial variation within a small group of pixels, the context of a pixel (or a group of pixels) refers to its spatial relationship with the local and global configuration of neighbouring pixels [Gurney and Townshedn, 1983; Binaghi *et al.,* 2003]. In VHR imagery, adjacent pixels are related or correlated [Khedama and Belhadj-Aissaa, 2004]. The spatial correlation or dependency arises due to the fact that spatial

resolution of the sensor is finer than the size of objects being classified [Shekhar *et al.,* 2002]. Information from neighbouring pixels (contextual information) plays an important role in the identification and extraction of urban features from remote sensing imagery [Jin and Davis, 2005] by increasing the discrimination capabilities of spectral pixel-based measured data. Moreover, in object-based approaches, when multi-scale segmentation is utilized, over-segmentation occurs in the lower levels of segmentation, resulting in too many boundaries such that real objects such as buildings and roads are split into two or more smaller objects. Hence, the spectral information of the object of interest is correlated with that of adjacent objects in the same level or with the spectral information of the super objects from the upper level. This contextual information can be incorporated into classification compensating for the spectral confusion between spectrally similar objects.

To date, little research has been conducted on incorporating contextual information in classification and object extraction of VHR imagery (compared to textural and morphological information). Table 2.4 lists the papers which have utilized contextual measures in the pixel-based or object-based classification of VHR imagery of urban areas.

Table 2.4 List of papers which have utilized contextual measures in pixel-based or object-based urban land cover classification of VHR imagery

Pixel-based classification	Object-based classification
Gurney and Townshed (1983); Binaghi *et al.,* (2003); Melgani and Sebastiano (2003); Shakelford and Davis (2003a); Shakelford and Davis (2003b); Khedam and Belhadj (2004); Jin and Davis (2005); ; Bellens *et al.,* (2008a); Miller *et al.,* (2009).	Herold *et al.,* (2003a); Shakelford and Davis (2003a); Thomas *et al.,* (2003); Nghi and Mai (2008) ; Chan *et al.,* (2009); Hermosilla *et al.,* (2011) ;

Shadows of high-rise features such as buildings and trees are among the widely used contextual measures. Several studies have utilized it in the pixel-based [e.g., Jin and Davis 2005; Bellens *et al.,* 2008b] or in the object-based [e.g., Shackelford and Davis 2003a; Nghi and Mai, 2008] classification processes in order to separate buildings from roads and streets in urban areas. Structural measures such as the length and width of a connected group of spectrally similar pixels are considered as contextual measures in pixel-based image analysis methods, while they are categorized as morphological measures in object-based methods. In the former case, these measures are calculated based on the spatial relationship between neighbouring pixels whereas in the latter case they are directly related to shape and size of each individual object. For example, Shackelford and Davis [2003b] used the length and width of a connected group of pixels, calculated based on spatial relationships within that group of pixels as two additional contextual bands in the classification. The length and width bands have high value for road and building classes, respectively.

Nghi and Mai [2008] utilized contextual relation in object-based classification. The contextual relation of an object is the number of objects that are adjacent with the object of interest. For instance, road objects have stronger relation than building objects. Thus, the relation can be used as additional information in the object-based classification to separate roads from buildings or shadows from large water bodies. In object-based approaches, the difference between the mean brightness value of the object of interest to its neighbour objects and to super objects for different bands are other types of contextual measures [Thomas *et al.,* 2003].Table 2.5 presents the results of some literature that have

used contextual information as additional bands in the urban land cover classification process of VHR imagery.

Table 2.5 Examples of using spatial measures (mainly contextual measures) for improving the urban land cover classification of VHR imagery. The spatial measures are incorporated as additional band into the classification. Accuracy improvement (Imp) is over the case where only spectral band(s) of the image is utilized in the classification process

Input data	Land cover types	Spatial measures	Classification approach	Accuracy	Paper
VHR-IK (Pansharp bands)	Road, building, tree, water, grass, shadow, bare soil	Texture: 1st order ENT for grass-tree Context: Length-width of a connected group of pixels for road-building and water-shadow	Pixel-based hierarchical fuzzy classification	OA: 92.7%,Imp:11% Road: PA:88%,Imp:17 % Building: PA:84%, Imp:11%	Shackelford and Davis (2003b)
VHR-IK (Pansharp bands)	Road, building, tree, water, grass, shadow, bare soil, impervious surf.	Context: Shadow of the objects Morphology: Skeleton of the objects	Pixel/Object-based Fuzzy classification	Building: PA:76%, Imp:3%, Road: PA:99%Imp:28 %	Shackelford and Davis (2003a)
VHR-QB	Water, grass, tree, roof(dark, red, white) road, other man made, shadow	Context: Shadow Proximity Feature(SPF), Shadow Distance Feature(SDF)	Pixel-based Maximum Likelihood	OA: 72% Imp:4% .In the case where only one context, either SPF or SDF is used	Bellens *et al.,* (2008a)
VHR-IK(Pan band)	Agricultural land, road networks, industrial plants, quarries, urban components (large buildings , small houses)	Context: A Multi-window set (6x6,20x20,56x56 pixels) of image data was used directly as the input of contextual information for MLP classifier	Pixel-based Multi Layer Perceptron (MLP) Neural Network	Correlation coefficient†:0.77 Imp: 0.28 Standard error:25.1 Imp: 28.8	Binaghi *et al.,* (2003)

*Feature Analyst software was released by Visual Learning System (VLS) in 2001and later on was developed as an extension for ESRI's Arc GIS and ERDAS Imaging (Miller et al., 2009).
†Accuracy assessments are based on regression analysis used for soft classification. Improvement is over the case when only a single window of 3x3, as oppose to multi-window set, was used with the same classification strategy.

2.2.3.3 Morphological Measures

The effectiveness of the aforementioned spatial measures, particularly texture measures, is highly dependent on the choice of optimal window size in pixel-based approaches and the level of segmentation in object-based ones. The optimal window size differs for

different classes in an urban scene, but defining an optimal window size or segmentation level for extracting textural and contextual measures is not an easy task. Moreover, members of same class can have different spectral reflectance values, such as black and gray building roofs [Miller *et al.,* 2009]. Consequently, they may have different textural and contextual measures. On the other hand, buildings, roads and parking lots possess specific morphological characteristics such as shape (smoothness and compactness) and size (length, width, perimeter, area, etc), especially in VHR images. A list of morphological measures can be found in the literature [e.g., Pratt, 2007 ; Definiens Developer 2007]. Incorporation of these morphological features into classification compensates for the lack of spectral information of VHR images and facilitates the discrimination process of spectrally similar classes. Subsequently, the classification accuracy of such classes increases.

Incorporating morphological measures in the pixel-based classification is usually done by applying a morphological filter to the image. Having applied the filter to the image, the spatial form or structure of objects within the image is modified. This modified image, then, is used as an additional band to the original bands of the image in classification. Dilation, erosion and skeletonization are three fundamental morphological operations [Pratt, 2007]. However, opening and closing, which are composed by combination of erosion and dilation operations, are the most common morphological operators used in literature in order to modify the form and structure of the objects within the image [Shackelford and Davis, 2003b]. Morphological measures are more meaningful and applicable in object-based classification approaches [Bellens *et al.,* 2008b] and many

researchers have benefited from their use. Table 2.6 lists the papers in which morphological measures have been utilized in pixel-based or object-based classification of VHR imagery of urban areas.

In object-based approaches the shape and size, as morphological measures of segmented regions, are directly utilized in classification. Table 2.7 as well as Table 2.5 report some examples of the use of morphological measures along with the classification approach and their corresponding accuracies in the urban land cover classification of VHR imagery.

Table 2.6 List of papers which have utilized morphological measures in pixel-based or object-based urban land cover classification of VHR imagery

Pixel-based classification	Object-based classification
Benediktsson *et al.,* (2003); Cablk and Minor (2003); Jin and Davis (2005); Mena and Maplica (2005); Bruzzone and Carlin (2006); Inglada (2007); Bellens *et al.,* (2008b); Fauvel *et al.,* (2008); Tuia *et al.,* (2009); *Chan et al* (2009); Lhomme *et al.(*2009).	Bauer and Steinnocher (2001);Herold *et al.,* (2003a); Shakelford and Davis (2003a); Thomas *et al.,* (2003); Carleer and Wolf (2006); Nghi and Mai (2008); Walker and Blashke (2008); van der Werff and van der Meer (2008); Chen *et al.,* (2009a); Zhou *et al.,* (2009); Xu and Li (2010); Hermosilla *et al.,* (2011); Pu *et al.,* (2011); Salehi *et al.,* (2011b); Salehi *et al.,* (2011c).

Table 2.7 Examples of using morphological measures (with their corresponding optimal factors) for improving urban land cover classification of VHR imagery. The morphological measures are incorporated as additional band into the classification. Accuracy improvement (Imp) is over the case where only spectral band(s) of the image is utilized in the classification process.

Input data	Land cover types	Morphological measures	Classification approach	Accuracy	Paper
VHR-QB (Full band)	Building and Transportation area	Length, width, length/width, area excluding inner regions, area including inner regions, perimeter, compactness	Object-based Nearest Neighbour	KC: Urban area:0.51 Suburban area:0.71	Carleer and Wolff (2006)
VHR- QB (Pan band)	Building, apartment blocks, road, railway, veg. ,trees, bare soil, soil, tower	Morphological operators: Opening, closing, reconstructed opening, reconstructed closing, with SE of the size of 9-25 pixels	Pixel-based SVM	OA:86.5%,Imp:52.5 % Build.:89%,Imp:62% Road:89%,Imp:35%	Tuia *et al.,* (2009)
VHR-QB (Pan and Pansharp)	Gray roof, Red roof, Road, Shadow, Rural area , Grass, Tree	From the 6th level of segmented image: Width/length ratio, Shape index, Rectangular fit	Pixel-based SVM classification of segmented image	OA:92.6% Imp:4%	Bruzzone and Carlin (2006)
VHR-IK (Full band)	Water ,grass, tree, buildings (dark, red and white roof), road, shadow, other man-made objects	Geometric Activity(GA) features including Ridge features based on facet model and Morphological features(closing)	Pixel-based Multi Layer Perceptron Neural Network	OA:71%,Imp*:5.4% Average PA of Man-made objects:72.2%, Imp*:5.3%	Chan *et al.,* (2009)
VHR-QB (Full band)	Bare soil ,grass, tree, roof, road	Morphological operators: Eight disk shaped SEs and eight line shaped SEs morphological profiles(MP) with partial reconstruction	Pixel-based Multilayer Perceptron Neural Network	OA:89%, Imp*:9% Road:72%, Imp:18.4% Roof:93%, Imp:17.6%	Bellens *et al.,* (2008b)
VHR-QB (Pansharp bands)	Building, road, tree, grass, soil, shadow, other impervious surface	7 Hu's, invariant moments 10 Zernike invariant moments 17 Wavelet invariant moments	Object-based SVM	OA: 80.5%, Imp‡:6.8%, Build:74.6%,Imp:6.4 % OA: 79.5%, Imp:5.8% Build:70.2%,Imp:2% OA: 80.2%, Imp: 6.5% Build: 70.2%,Imp:2%	Xu and Li (2010)

*Improvement is over pixel-based MLP neural network when only spectral information (i.e., 4 MS bands + Pan and NDVI bands) is used

†Improvement is over the same pixel-based neural network when only 1st component of 40 principal components of the original data is used

‡Improvement is over the same object-based SVM when only spectral information (four Pan-sharp bands) is used.

2.2.4 Spatial Measures Extracted from the Ancillary Data

Spatial information can be derived from the image itself (e.g., texture, context and morphology), which was broadly discussed in previous sections and/or from other data sources, the so-called ancillary data. Ancillary data layers are key components of accurate image classification [Thomas *et al.,* 2003]. Particularly, with the widespread availability of VHR imagery, digital elevation/surface model (DEM/DSM) extracted from LiDAR data or stereo images and existing GIS data layers, the importance of integrating these data for detailed mapping of urban environments becomes highly significant. Heinl *et al.,* [2009] showed that the use of ancillary data improves the classification accuracy independent of classification method.

When ancillary data are utilized along with the image in the classification process, usually the misregistration between the ancillary data and the image is a problematic issue. Precise geometric registration of corresponding data layers is often very difficult to achieve, particularly for VHR imagery. Since the basic mapping unit in object-based approaches is a group of connected pixels (i.e., object instead of pixel), the misregistration between multisource data (i.e., VHR image, LiDAR and GIS data layers) is not as serious as for pixel-based approaches [Justice *et al.,* 1989; Zhou *et al.,* 2008]. In fact, object-based approaches facilitate the use of ancillary data [Kim *et al.,* 2010] and since they require less precise registration of the image, they are highly desirable for multisource image analysis [Kim *et al.,* 2010].

Several methods may be used to incorporate ancillary data into the classification process. The most common used is the stacked vector or the logical channel method [Jensen 2005,

Watanachaturaporn *et al.,* 2008; Huang *et al.,* 2011], which considers the ancillary data as an extra channel (band) to the original channels of the image in pixel-based or object-based classification . In object-based approaches, in addition to the stacked layer method, ancillary data are also used in a rule-based manner during segmentation [e.g., Bouziani *et al.,* 2010] and classification.

Two groups of ancillary data are frequently used: DSM produced from LiDAR (LiDAR-derived DSM) data or from aerial or satellite stereo images (photogrammetrically-derived DSM) and GIS data layers such as the map of parcels, the road centreline network, etc. Table 2.8 lists the papers in which ancillary data have been incorporated into classification (both pixel-based and object-based) of VHR imagery of urban areas. As seen in this table, object-based approaches have utilized ancillary data significantly more than have pixel-based approaches.

Table 2.8 List of papers which have utilized ancillary data in the pixel-based or object-based urban land cover classification of VHR imagery

	Pixel-based classification	**Object-based classification**
Height (Mainly LiDAR data)	Hodgson *et al.,* (2003); Rottensteiner *et al.,* (2003); Rottensteiner *et al.,* (2005); Huang *et al.,* (2011).	Schiewe and Ehler (2005); Brennan and Webster (2006); Zhou and Troy (2008); Zhou *et al.,* (2008); Aubrecht *et al.,* (2009); Chen *et al.,* (2009b); Zhou *et al.,* (2009); Hussain and Shan (2010); Yu *et al.,* (2010); Hermosilla *et al.,* (2011); Salehi *et al.,* (2011b);Wurm *et al.,* (2011).
GIS data layers	Thomas *et al.,* (2003)	Sims and Mesev (2007); Zhou and Troy (2008); Zhou *et al.,* (2008); Aubrecht *et al.,* (2009); Bouziani *et al.,* (2010).

2.2.4.1 DSM Derived from LiDAR and Stereo Photographs

DSM generated by LiDAR data or by stereo aerial/satellite images (photogrammetrically-derived DSM) gives information about the height of objects (e.g., buildings and trees); thus it is very helpful in separating spectrally similar objects with different heights (e.g.,

buildings from streets and tree from shadows). LiDAR- and photogrammetrically-derived DSM represent the height of each pixel with respect to the reference datum. Consequently, when they are intent to be incorporated in the classification process, the first stage involves removal of the underlying terrain, so-called digital elevation model (DEM), from the DSM to produce the heights of objects above the local ground (e.g., building's height) [Ma 2005].The resultant height is referred to as normalized DSM (nDSM) [Chen *et al.,* 2009b*]*. The nDSM could then be integrated along with the spectral bands of the image as an additional channel in both segmentation and classification processes or as additional information in rule-based segmentation and classification. The influence of DSM on classification can be controlled by adjusting channels' weights in the former method [Hofmann, 2001].

There is a considerable amount of research accomplished in the integration of LiDAR DSM with the image during classification, particularly in object-based analysis of VHR images over urban areas (Table 2.8). Regardless of the classification approach, results show a significant improvement of classification accuracy [e.g., Sohn and Dowman, 2007; Watanachaturaporn *et al.,* 2008; Huang *et al.,* 2011].Very few studies, however, have benefited from the incorporation of the photogrammetrically-derived DSM and VHR imagery for classification over urban areas. This is mainly due to unavailability of precise DSM of urban areas and misregistration between the DSM and VHR imagery. However, recent development of stereo satellite and aerial VHR imagery in conjunction with advancements in object-based image analysis methods has facilitated the integration of the photogrammetrically-derived DSM and VHR data.

Hussain and Shan [2010] integrated photogrammetrically-derived nDSM and VHR imagery in a rule-based object-based classification method in order to separate buildings from transportation areas and the result was very satisfactory [Hussain and Shan, 2010]. They further employed the nDSM to separate different buildings according to their heights (i.e. single and double story houses and apartments). Table 2.9 presents a summary of some recent papers in which ancillary data (mainly LiDAR nDSM) have been utilized together with spectral bands of the VHR imagery for object-based and pixel-based classification.

Table 2.9 Examples of using ancillary data (mainly LiDAR nDSM) in addition to VHR imagery for improving urban land cover classification. The ancillary measures are incorporated as additional band into the classification and/or segmentation. Accuracy improvement (Imp) is over the case where only spectral band(s) of the image is utilized in the classification process.

Input data	Ancillary data used	Land cover types	Classification approach	Accuracy	Paper
VHR-QB Pansharp bands	LiDAR nDSM: HA*=30cm, VA†=15cm SR=1m	Water, shadow, grass, shrub, building, road, vacant areas	Hierarchical object-based	OA:89.4%,Imp:20% Building:PA:92.5%, Imp:9% Road:PA:86%, Imp,47.5%	Chen *et al.,* (2009b)
VHR-Aerial image (0.6m SR in G,R, NIR bands)	LiDAR nDSM: First and last pulse data, SR=1m, GIS data: Parcel boundary, Building footprints	Building, pavement, coarse texture vegetation, fine texture vegetation, bare soil	Hierarchical rule-based object-oriented	OA: 92.3% Building PA:94.4% Pavement PA: 88.3% Imp: NM	Zhou and Troy (2008)
LiDAR intensity	LiDAR nDSM, DSM, Echo code (Multiple return)	Water, low veg., road structure, deciduous, coniferous, intertidal	Hierarchical rule-based object-oriented	OA: 98.1%, Structure PA:94.2% Road PA:94.5% Imp: NM	Brennan and Webster (2006)
VHR-Aerial orthophoto (0.4 m in R,G,B bands)	LiDAR DSM HA: 0.5 m, VA: 0. 15 m A Max-Min band resulted from moving a window of 13x13 size over LiDAR DSM	Ground, grass, shadow, building, tree	Pixel-based SVM	OA: 94.7%, Imp‡: 12.2% Build: 96.4% Imp: 29.6%	Huang *et al.,* (2011)

* HA: Horizontal Accuracy; † VA: Vertical Accuracy

‡ Improvement is over pixel-based SVM when only spectral information (3 multispectral bands) is used.

2.2.4.2 GIS Data Layers

Despite the growing attention toward use of LiDAR data for classification,few studies have taken the advantages of existing GIS data layers for improving urban land cover classification (Table 2.8). Thomas *et al.,* [2003] developed strategies including spatial modelling techniques to deal with the problem of confusion between spectrally similar classes in VHR data. Their spatial modelling was based on the integration of GIS ancillary data layers with the image bands. They used the distance from road centrelines to differentiate buildings and parking lots from roads Zhou and Troy [2008] employed building footprint data in a rule-based hierarchical object-oriented classification to separate buildings from non-building objects. They showed that the classification accuracy increased when building footprint data together with LiDAR DSM are incorporated into the classification process.

2.3 Discussion

2.3.1 Spectral Measures

The result of this literature review shows that spectral information of VHR imagery plays a major role in classifying vegetation (including grass and trees), water surfaces and even shadow areas. In particular, pixel-based approaches give superior results over object-based classification for mapping such land cover types in a typical urban environment. However, impervious land cover types (e.g., buildings, roads and parking areas) are spectrally too similar to be separated only according to their spectral information in VHR imagery. Between pixel-based and object-based classification approaches, a serious

problem common to both approaches arises in the classification of spectrally similar and heterogeneous classes.

2.3.2 Spatial Measures

2.3.2.1 Pixel-based vs. Object-based Classification

Figure 2.2 presents the number of papers in which spatial measures have been utilized in pixel-based or object-based classification approaches. This figure is a summary of Tables 2.1, 2.4, 2.6, and 2.8. Texture features have been utilized in pixel-based approaches significantly more than in object-based approaches. On the other hand, ancillary data, especially LiDAR DEM, have been used in object-based classification more often than in pixel-based classification. The number of papers that have used morphological measures is almost equal in both pixel-based and object-based classification (with two papers more in object-based). Ten papers have used contextual measure in pixel-based classification and seven papers used them in object-based approaches. Compared to the other measures, GIS data have been employed in a very few papers (five object-based and one pixel-based) indicating the very early stage of combining GIS data and VHR imagery for classification purposes.

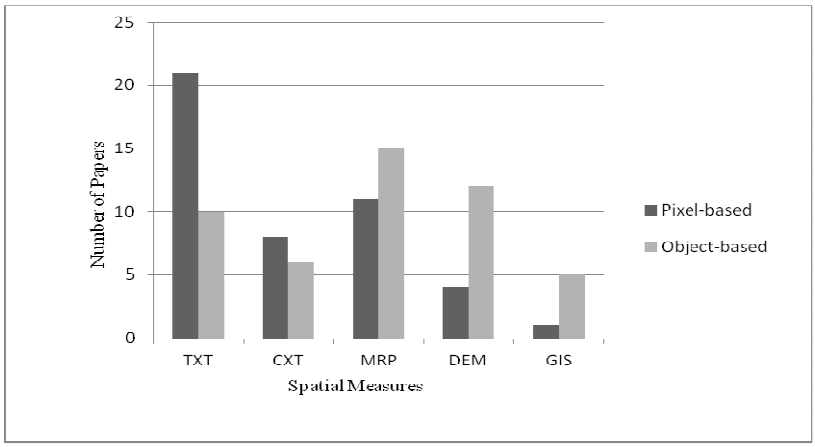


Figure 2.2 The number of papers which have utilized spatial measures (TXT:texture, CXT: context, MRP:morphology, DEM, height data, and GIS:GIS data) in pixel-based and object-based classification of VHR imagery over urban areas.

2.3.2.2 Textural Measures

From the accuracies reported in the literature, texture has a significant effect on improving the classification accuracy of urban areas using VHR imagery. The improvement ranges from 2% to11% in terms of overall accuracy. Some of the major findings on the performance of texture measures are as follows:

- Texture measure, in general, has better performance in pixel-based than object-based approaches.
- Texture does not necessarily increase the classification accuracy for all classes. It is desirable to incorporate texture measures in classification using a fuzzy-based classification with different membership for different sets of classes.
- For urban impervious surfaces, integrating texture increases the classification accuracy, whereas for vegetation and water, texture does not have significant effect or even lowers the classification accuracy.

- In general larger window sizes of texture work better in classifying heterogeneous land cover types for VHR imagery, whereas smaller window size is preferred for lower resolution imagery.
- In object-based image analysis, when multilevel segmentation is used, texture measures of objects at higher levels of segmentation would be more appropriate for classification of heterogeneous land cover types such as impervious surfaces, while for homogeneous classes, texture measures of objects at lower levels of segmentation are preferred.

2.3.2.3 Contextual Measures

The amount of literature in which contextual information is utilized in the classification process is relatively small compared to that using texture and morphology (Figure 2.2). Shadow is the most used contextual measure in classification, especially in pixel-based methods. The spectral and spatial relation between objects in the same or different levels of segmentation is the major source of contextual measures in object-based methods. Furthermore, contextual measures are rarely used as the only spatial measure in classification. In other words, contextual measures are often used in conjunction with textural and/or morphological measures during classification. Nevertheless, in all cases where contextual measures have been incorporated, classification accuracies have been increased.

2.3.2.4 Morphological Measures

The role of morphological information in object-based classification is more significant than in pixel-based. Indeed, a key advantage of object-based over pixel-based image analysis is that the results of segmentation in object-based approaches are sets of meaningful regions for which a number of morphological features can be directly measured and used during the classification process. Moreover, some segmentation algorithms are able to create objects in different scales with different sizes and shapes. This multi scale or multi level image segmentation allows the classifier to utilize morphological properties of objects in various scales, resulting in higher classification accuracy for classes such as roads and buildings, which present in different sizes and shapes. The quantitative results of some studies showed an average accuracy of around 90% in terms of producer's accuracy (PA) for buildings and roads in a typical urban environment, when both spectral and morphological measures of VHR imagery are utilized in object-based classification.

2.3.2.5 DSM Derived from LiDAR and Stereo Photographs

Recent developments in object-based image analysis and increasingly available LiDAR data and VHR imagery have directed researchers' attention toward their integration for classification purposes. Nevertheless, we learned that the use of LiDAR data along with VHR imagery for detailed land cover classification is still at an early stage of development, although they have exhibited potential in urban land cover mapping. Many of the published papers in recent years are conference papers and the results show that the

incorporation of LiDAR data into the classification of VHR imagery can significantly solve the problem of differentiating high-rise and low-rise objects with similar spectral reflectance.

For the case of integration of VHR imagery and photogrammetrically-derived DSM, very few papers have been found. With the availability of stereo VHR imagery from satellites such as QuickBird, GeoEye-1 and WorldView-2, the advances in object-based image analysis, and precise DSM generation methods, the integration of VHR imagery and its photogrammetrically-derived DSM for mapping of complex urban landscapes is more feasible than before. It is worth mentioning that it is important to avoid misregistration problems between DSMs and VHR images. Almost all papers avoided the use of off-nadir VHR images in the classification. In the real world, however, more than 90% of VHR images are collected off nadir.

2.3.2.6 GIS Data Layers

Traditionally, existing GIS map layers such as road /street networks and building footprints have been used as reference data to evaluate the performance of classification. Recently, the potential of utilizing this vector data for improving classification accuracy has drawn increased attention from many researchers. Because GIS data layers are somewhat consistent with object-based classification input, i.e., object resulting from segmentation may share the same boundaries with GIS data, these data have been increasingly used to improve object-based classification. Nonetheless, the use of GIS data in classification, in the literature, is far less than that of other spatial measures (Figure 2.2). Misregistration between GIS vectors and images still poses a great challenge for

integration. Most papers used either low resolution images or nadir VHR images in the classification to avoid the errors introduced by misregistration.

2.4 Conclusion

Over the past decade, there have been ever-growing numbers of researchers studying on detailed land cover classification of urban areas. This is partly due to the fact that the differentiation amongst three major impervious classes in an urban scene (roads, buildings and parking lots) becomes more feasible with the availability of VHR imagery. Due to the complex nature of urban landscapes as well as the spatial and spectral characteristics of VHR imagery, the classification of such landscapes requires not only spectral but also spatial information of the image. For this reason, the amount of literature using spatial information includes texture, context, morphology and information extracted from ancillary data such as DSM and archived GIS layers has grown since the launch of first VHR satellite in 1999.

Although spatial measures have been used in both pixel-based and object-based classification approaches, the employment of them in the latter case is more effective mainly due the following reasons: First, determining the optimal window size of spatial measures, which is a critical issue in pixel-based classification methods, has been solved in object-based approaches by segmenting the image to individual objects with meaningful boundaries. Second, contextual and especially morphological measures are more meaningful in the object-based image analysis. Third, object-based approaches require less precise geometric registration between different data layers when ancillary

data are employed in the classification. However, the major drawback of object-based image classification is the lack of a high degree of automation. In fact, different rule-sets must be developed for different image and applications.

Among the three spatial measures extracted from the images, texture has most often employed in classification, especially in pixel-based approaches. Multiresolution segmentation of object-based approaches, on the other hand, enhances the capability of morphological measures to improve the classification of land covers such as buildings and traffic areas. The classification accuracies reported in literature indicate that morphological measures have significantly higher effect on differentiating classes such as buildings and roads than textural and contextual measures. The use of contextual measures in classification of VHR imagery is less than the use of textural and morphological measures. In addition, contextual measures have rarely been used as the only source of spatial measure in classification. Indeed, the effectiveness of such measures is enhanced when they are used in conjunction with other spatial measures (textural and/or morphological measures).

The employment of LiDAR- and especially photogrammetrically-derived DSM in urban land cover classification of VHR imagery is in the very early stages perhaps because of non availability of LiDAR data and the rather low precision of photogrammetrically-derived DSM. Nonetheless, the results of our literature review reveal the very high potential of this type of data in conjunction with VHR imagery for land cover mapping of urban environments. With the widespread availability of VHR stereo aerial/satellite imagery and the development in object-based image analysis and precise DSM

generation methods, the use of photogrammetrically-derived DSM in classification of VHR imagery over urban areas is more feasible than before and has a very high potential for future research. Also despite the availability of archived GIS data layers, this type of ancillary data has not been well utilized for classification purposes. The development of spatial modelling of available GIS data layers for incorporation into the classification process, particularly object-based classification approaches, would be another interesting topic for future research.

Acknowledgements

This research is supported by the NSERC (Natural Science and Engineering Research Council of Canada) Strategic Project Grants and NSERC Discovery Grants Programs.

REFERENCES

Aksoy, S., Koperski, K.,Tusk,C., and Marchisio, G. 2009. Land cover classification with multi-sensor fusion of partly missing data. *Photogrammetric Engineering and Remote Sensing*, Vol. 75, No. 5, pp. 577-598.

Agüera, F., Aguilar, F.J. and Aguilar, M.A. 2008. Using texture analysis to improve per-pixel classification of very high resolution images for mapping plastic greenhouses. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 63, No. 6, pp. 635-646.

Alonso, M.C., Sanz, M.A. and Malpica, J.A. 2007. Classification of high resolution satellite images using texture from the panchromatic band. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4842 LNCS, Part 2, pp.499-508.

Anys, H., Bannari, A., He, D.C. and Morin, D. 1994. Texture analysis for the mapping of urban areas using airborne Meis-II images. In *Proceeding* of *the First International Airborne Remote Sensing Conference and Exhibition: Applications, Technology, and Science,* Strasbourg, France, Vol. 3, pp. 231-245.

Aubrecht C., Steinnocher, K.,Hollaus, M. and Wagner, W. 2009. Integrating earth observation and GIScience for high resolution spatial and functional modeling of urban land use. *Computers, Environment and Urban Systems*, Vol. 33, No. 1, pp. 15–25.

Bauer, T. and Steinnocher, K. 2001. Per-parcel land use classification in urban areas applying a rule-based technique. *Geo-Informations-Systeme,* Vol. 14, No. 6, pp. 24-27.

Bellens, R., Douterloigne, K., Gautama, S. and Philips, W. 2008a. Per pixel contextual information for classification of VHR images of urban areas. In *IEEE International Geoscience and Remote Sensing Symposium,* Boston.

Bellens, R., Gautama, S., Martinez-Fonte, L., Philips, W., Chan, J.C. and Canters, F. 2008b. Improved Classification of VHR Images of Urban Areas Using Directional Morphological Profiles. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 46, No. 10, pp. 2803-2813.

Ben-Dor, E. 2001. Imaging Spectroscopy for Urban Applications in *Imaging Spectroscopy* F. Van Der Meer and S.M. de Jong (Eds), pp. 243-281(Kluwer Academic Press).

Benediktson, J.A., Pesaresi, M. and Arnason, K. 2003. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 41, No. 9(Part1), pp. 1940-1949.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I. and Heynen, M. 2004. Multiresolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 58, No. 3-4, pp. 239-258.

Binaghi, E., Gallo, I. and Pepe, M. 2003. A cognitive pyramid for contextual classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 4, No. 12, pp. 2906-2922.

Blaschke, T. 2010. Object-based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 65, No. 1, pp. 2-16.

Blaschke, T., Lang, S. and Hay, G.J. 2008. *Object-Based Image Analysis: Spatial concepts for Knowledge-Driven Remote Sensing Applications. Se*ries: XVII, *Lecture Notes in Geoinformation and Cartography* (Berlin: Springer-Verlarge).

Bouziani, M.,Goita, K. and He, D.C. 2010. Rule-based classification of a very high resolution image in an urban environment using multispectral segmentation guided by cartographic data. IEEE Transactions on Geoscience and Remote Sensing, Vol. 48, No. 8, pp. 3198-3211.

Brennan, R., Webster, T.L. 2006. Object-oriented land cover classification of lidar-derived surfaces. *Canadian Journal of Remote Sensing*, Vol. 32, No. 32, pp. 162-172.

Bruzzone, L. and Carlin, L. 2006. A multilevel context-based system for classification of very high spatial resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44, No. 9, pp. 2587-2600.

Cablk, M.E. and Minor, T.B. 2003. Detecting and discriminating impervious cover with high-resolution IKONOS data using principal component analysis and morphological operators. *International Journal of Remote Sensing*, Vol. 24, No. 23, pp. 4627-4645.

Carleer, A.P. and Wolff, E. 2006. Urban land cover multi-level region-based classification of VHR data by selecting relevant features. *International Journal of Remote Sensing*, Vol. 27, No. 6, pp. 1035-1051.

Chan, J.C., Bellens, R., Canters, F. and Gautama, S. 2009. An assessment of geometric activity features for per-pixel classification of urban man-made objects using very high resolution satellite imagery. *Photogrammetric Engineering and Remote Sensing*, Vol. 75, No. 4, pp. 397-411.

Chen, D., Stow, D.A. and Gong, P. 2004. Examining the effect of spatial resolution and texture window size on classification accuracy: An urban environment case. *International Journal of Remote Sensing*, Vol. 25, No. 11, pp. 2177-2192.

Chen, M., Su, W., Li, L., Zhang, C., Yue, A. and Li, H. 2009a. Comparison of pixel-based and object-oriented knowledge-based classification methods using SPOT5 imagery. *WSEAS Transactions on Information Science and Applications,* Vol. 6, No. 3, pp. 477-489.

Chen, Y., Su, W., Li, J. and Sun, Z. 2009b. Hierarchical object-oriented classification using very high resolution imagery and LIDAR data over urban areas. *Advances in Space Research*, Vol. 43, No. 7, pp. 1101-1110.

Congalton, R.G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, Vol. 37, No. 1, pp. 35-46.

Cossu, R. 1988. Segmentation by means of textural analysis. *Pixel* , Vol. 1, No. 2, pp. 21-24.

Definiens, 2007. *eCognition Professional User Guide*. Avialable online at : http://www.definiens.com (accessed March 2011).

Fauvel, M., Bnediktsson, J. A., Chanussot, J. and Sveinsson, J. R. 2008. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 46, No. 11, pp. 3804-3814.

Gamba, P., Dell'Acqua, F. and Dasarathy, B.V. 2005. Urban remote sensing using multiple data sets: Past, present, and future. *Information Fusion*, Vol. 6, No. 4, pp. 319-326.

GEOBIA, 2008. *The Proceedings of GEOgraphic Object Based Image Analysis for the 21 Century*. Calgary, Canada. Available online at: http://homepages.ucalgary.ca/~gjhay/geobia/Proceedings /Sessions/ GEOBIA_Proceedings_RP10_FINALc.pdf (accessed March 2011).

GEOBIA , 2010. *The Proceedings of GEOgraphic Object Based Image Analysis.* Ghent, Belgium 29 June-2July, 2010 Available online at: http://geobia.ugent.be/ proceedings/html/papers.html (accessed March, 2011).

Gong, P., Marceau, D.J. and Howarth, P.J. 1992. A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data. *Remote Sensing of Environment,* Vol. 40, No. 2, pp. 137-151.

Gurney, C. M. and Townshend, J. R. G. 1983. The use of contextual information in the classification of remotely sensed data. *Photogrammtric Engineering and Remote Sensing,* Vol. 49, No. 1, pp. 55–64.

Haralick, R.M. 1979. Statistical and structural approaches to texture. *In Proceedings of IEEE* , Vol. 67, No. 5, pp. 786-804.

Heinl, M., Walde, J., Tappeiner, G. and Tappeiner, U. 2009. Classifiers vs. input variables-the drivers in image classification for land cover mapping. International Journal of Applied Earth Observation and Geoinformation, Vol. 11, No. 6, pp. 423-430.

Herold, M., Liu, X. and Clarke, K.C. 2003a. Spatial metrics and image texture for mapping urban land use. *Photogrammtric Engineering and Remote Sensing,* 69, pp. 991–1001.

Herold, M., Gardner, M.E. and Roberts, D.A. 2003b. Spectral resolution requirements for mapping urban areas. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 41, No.9 (Part 1), pp. 1907-1919.

Herold, M., Roberts, D.A., Gardner, M.E. and Dennison, P.E. 2004. Spectrometry for urban area remote sensing-Developement and analysis of a spectral library from 350 to 2400 nm. *Remote Sensing of Environment*, Vol. 91, No. 2/4, pp. 304-319.

Hermosilla, T., Ruiz, L. A.Recio, J. A. and Estornell, J. 2011. Evaluation of Automatic Building Detection Approaches Combining High Resolution Images and LiDAR Data. *Remote Sensing,* Vol. 3*,* No. 6, pp. 1188-1210.

Hester, D.B., Cakir, H.I., Nelson, S.A.C. and Khorram, S. 2008. Per-pixel classification of high spatial resolution satellite imagery for urban land-cover mapping. *Photogrammetric Engineering and Remote Sensing*, Vol. 74, No. 4, pp.463-471.

Hodgson, M.E., Jensen, J.R., Tullis, J.A., Riordan, K.D. and Archer, C.M. 2003. Synergistic use of lidar and color aerial photography for mapping urban parcel imperviousness. *Photogrammetric Engineering and Remote Sensing,* Vol. 69, No. 9, pp. 973-980.

Hofmann, P. 2001. Detecting buildings and roads from IKONOS data using additional elevation information. *Geo-Informations-Systeme,* Vo.14, No. 6, pp. 28-33.

Huang, X., Zhang, L. and Gong, W. 2011. Information fusion of aerial images and LIDAR data in urban areas: vector-stacking, re-classification and post-processing approaches. *International Journal of Remote Sensing*, Vol. 32, No. 1, pp. 69-84.

Hussain, E. and Shan, J. 2010. Rule inheritance in object-based image classification for urban land cover mapping. In: Proceedings of ASPRS Annual Conference, April 26-30, San Diego, California.

Inglada, J. 2007. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 62, No.3, pp. 236-248.

Jensen, J.R. 2005. *Introductory digital image processing: a remote sensing perspective,* 3rd ed. (Upper Saddle River, N.J.: Prentice Hall).

Jin, X. and Davis, C.H. 2005. Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *Eurasip Journal on Applied Signal Processing,* Vol. 14, pp. 2196-2206.

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,* Vol. 10, No. 9, pp. 1539-1561.

Khedama, R. and Belhadj-Aissaa, A. 2004. Contextual classification of remotely sensed data using MAP approach and MRF. In *Proceeding of ISPRS 2004 Congress*, Istanbul, Turkey.

Kim, M., Madden, M. and Xu, B. 2010. GEOBIA vegetation mapping in general smoky mountains national park with spectral and non-spectral ancillary information. *Photogrammetric Engineering and Remote Sensing*, Vol. 76, No. 2, pp.137-148.

Liu, Y., Li, M., Mao, L., Xu, F. and Huang, S. 2006. Review of remotely sensed imagery classification patterns based on object-oriented image analysis. *Chinese Geographical Science*, Vol. 16, No. 3, pp. 282-288.

Lhomme, S.,He, D.C.,Weber, C. and Morin, D. 2009. A new approach to building identification from very high- spatial-resolution images. *International Journal of Remote Sensing*, Vol. 30, No. 5, 1341-1354.

Lu, D., Hetrick, S. and Moran, E. 2010. Land cover classification in a complex urban-rural landscape with Quickbird imagery. *Photogrammetric Engineering and Remote Sensing,* Vol. 76, No. 10, pp.1159-1168.

Lu, D. and Weng, Q. 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, Vol. 28, No. 5, pp. 823-870.

Luo, L., and Mountrakis, G. 2010. Integrating intermediate inputs from partially classified images within a hybrid classification framework: An impervious surface estimation example. *Remote Sensing of Environment,* Vol. 114, No. 6, pp. 1220-1229.

Ma, R. 2005. DEM generation and building detection from lidar data. *Photogrammetric Engineering and Remote Sensing,* Vol. 71, No. 7, pp. 847-854.

Maillard, P. 2003. Comparing texture analysis methods through classification. *Photogrammetric Engineering and Remote Sensing*, Vol. 69, No. 4, pp. 357-367.

McMahon, G. 2007. Consequences of land-cover misclassification in models of impervious surface. *Photogrammetric Engineering and Remote Sensing*, Vol. 73, No.12, pp.1343-1353.

Mena, J. B. and Malpica, J. A. 2005. An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery. *Pattern Recognition Letters*, Vol. 26, No. 9, pp. 1201-1220.

Melgani, F. and Serpico, S.B. 2003. A Markov Random Field approach to spatio-temporal contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing* Vol. 41, No. 11(Part 1) pp. 2478-2487.

Miller, J.E., Nelson, S.A.C. and Hess, G.R. 2009. An object extraction approach for impervious surface classification with very-high-resolution imagery. *Professional Geographer*, Vol. 61, No. 2, pp. 250-264.

Myint, S.W. 2007. Urban mapping with geospatial algorithms. In: *Urban remote sensing* Q.Weng, D.A Quattrochi (Eds.), pp. 109-135 (Boca Raton, Florida: Taylor & Francis).

Myint, S.W. and Lam, N.S. 2005. A study of lacunarity-based texture analysis approaches to improve urban image classification. *Computer, Environment and Urban System*, Vol. 29, No. 5 pp.501-523.

Myint, S.W., Lam, N.S. and Tyler, J.M. 2004. Wavelets for urban spatial feature discrimination: Comparisons with fractal, spatial autocorrelation and spatial co-occurrence approaches. *Photogrammetric Engineering and Remote Sensing,* Vol. 70, No. 7, pp. 803-812.

Nghi, D.L. and Mai, L.C. 2008. An object-oriented classification techniques for high resolution satellite imagery. In *the Proceedings of International Symposium on Geoinformatics for Spatial Infrastucture Development in Earth and Applied Science.*

OBIA, 2006. *The of Proceedings 1st International Conference on Object-based Image Analysis,* July 2006 Salzburg, Austria. Available online at: http://www.commission4.isprs.org/obia06/index.html (Accessed April, 2010).

Ouma, Y.O., R. Tateishi, R., and Sri-Sumantyo, J. T. 2010. Urban features recognition and extraction from very-high resolution multi-spectral satellite imagery: a micro–macro texture determination and integration framework. *IET Image Processing,* Vol. 4, No. 4, pp. 235-254

Pacifici, F., Chini, M. and Emery, W.J. 2009. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment*, Vol. 113, No. 6, pp. 1276-1292.

Pesaresi, M. 2000. Texture analysis for urban pattern recognition using fine-resolution panchromatic satellite imagery. *Geographical & Environmental Modelling,* Vol. 4, No.1, pp.43- 63.

Pratt, W.K. 2007. *Digital Image Processing*, 4th ed. Wiley.

Price, J.C. 1997. Spectral band selection for visible-near infrared remote sensing: spectral-spatial resolution tradeoffs. IEEE Transactions on Geoscience and Remote Sensing, Vol. 35, No. 5, pp. 1277-1285.

Pu, R. Landry, S. and Qian Y. 2011. Object-based urban detailed land cover classification with high spatial resolution IKONOS imagery. *International Journal of Remote Sensing,* Vol. 32, No. 12, pp. 3285-3308.

Puissant, A., Hirsch, J. and Weber, C. 2005. The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. *International Journal of Remote Sensing,* Vol. 26, No. 4, pp. 733-745.

Quattrochi, D.A. and Goodchild, M.F. 1997. *Scale in remote sensing and GIS* (Boca Raton, Florida :Lewis Publishers).

Richards, J.A., Jia, X. 2006. *Remote sensing digital image analysis : An introduction,* 4th ed. (Berlin: Springer).

Rottensteiner, F.,Trinder, J., Clode, S. and Kubik, K. 2003. Building detection using LiDAR data and multi-spectral images. *In Proc. of VIIth Digital Image Computing: Techniques and Applications*, Sydney, Australia.

Rottensteiner, F.,Summer, G., Trinder, J., Clode, S. and Kubik, K. 2005. Evaluation of a method for fusing Lidar data and multispectral images for building detection. In: Stilla U, Rottensteiner F, Hinz S (Eds) CMRT05. IAPRS, Vol. XXXVI, Part 3/W24 - Vienna, Austria.

Salehi, B., Zhang,Y. and Zhong, M. 2011a. Combination of object-based and pixel-based image analysis for the classification of VHR imagery over urban areas. *In the Proceedings of ASPRS 2011 Annual Conference*, May 1-5, Milwaukee, Wisconsin, USA.

Salehi, B., Zhang,Y. and Zhong, M. 2011b. Object-based land cover classification of urban areas using VHR imagery and photogrammetrically-derived DSM. *In the Proceedings of ASPRS 2011 Annual Conference,* May 1-5, Milwaukee, Wisconsin, USA.

Salehi, B. Zhang, Y. and Dey, V. 2011c. A hierarchical rule based urban land cover classification of WorldView-2 imagery, *In Proceedings of 32th Canadian Symposium on Remote Sensing*, Sherbrook, Canada.

Schiewe, J. and Elers, M. 2005. A novel method for generating 3D city models from high resolution and multi-sensor remote sensing data. *International Journal of Remote Sensing*, Vol. 26, No. 4, pp. 683-698.

Shackelford, A.K. and Davis, C.H. 2003a. 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*, Vol. 41, No. 10 (Part1), pp. 2354-2363.

Shackelford, A.K. and Davis, C.H. 2003b. A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, No. 9 (Part 1), pp.1920-1932.

Shekhar, S., Schrater, P.R., Vatsavai, R.R., Wu,W. and Chawla, S. 2002. Spatial contextual classification and prediction models for mining geospatial data. IEEE Transactions on Multimedia, Vol. 4, No. 2, pp. 174-188.

Sims, F.M. and Mesev, V. 2007. Use of ancillary data in object-based classification of high resolution satellite data.In *Proceedings of Urban Remote Sensing Joint Event*. April 11-13, 2007, Paris,France.

Sohn, G. and Dowman, I. 2007. Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 62, No. 1, pp. 43-63.

Song, M., Civco, D.L. and Hurd, J.D. 2005. A competitive pixel-object approach for land cover classification. *International Journal of Remote Sensing*, Vol. 26, No. 22, pp.4981-4997.

Su, W., Li, J., Chen, Y., Liu, Z., Zhang, J., Low, T.M., Suppaih, I. and Hashim, S.A.M. 2008. Textural and local spatial statistics for the object-oriented classification of urban areas using high resolution imagery. International Journal of Remote Sensing, Vol. 29, No. 11, pp. 3105-3117.

Tassetti, A. N., Malinverni, E.S., and Hahn, M. 2010. Texture analysis to improve supervised classification in Ikonos imagery. *In: Wagner W., Székely, B. (eds.): ISPRS TC VII Symposium – 100 Years ISPRS, Vienna, Austria, IAPRS, Vol. XXXVIII, Part 7A.*

Thomas, N., Hendrix, C. and Congalton, R.G. 2003. A comparison of urban mapping methods using high-resolution digital imagery. *Photogrammetric Engineering and Remote Sensing*, Vol. 69, No. 9, pp. 963-972.

Tuia, D., Pacifici, F., Kanevski, M. and Emery, W.J. 2009. Classification of Very High Spatial Resolution Imagery Using Mathematical Morphology and Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47, No. 11, pp. 3866-3879.

Van de Voorde, T., De Genst, W. and Canters, F. 2007. Improving pixel-based VHR land-cover classifications of urban areas with post-classification techniques. *Photogrammetric Engineering and Remote Sensing*, Vol. 73, No. 9, pp. 1017-1027.

Van der Werff , H.M.A. and Van der Meer F.D. 2008. Shape-based classification of spectrally identical objects. *ISPRS Journal of Photogrammetry & Remote Sensing,* Vol. 63, No. 2, pp. 251–258.

Walker, J. and Blaschke, T. 2008. Object-based land cover classification for the Phoenix metropolitan area: optimization vs. transportability. *International Journal of Remote Sensing,* 29, pp. 2021-2040.

Walter, V. 2004. Object-based classification of remote sensing data for change detection. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 58, No. 3-4, pp. 225– 238.

Warner, T.A. and Nerry, F. 2009. Does single broadband or multispectral thermal data add information for classification of visible, near and shortwave infrared imagery of urban areas? *International Journal of Remote Sensing*, Vol. 74, No. 7, pp. 2155-2171.

Watanachaturaporn, P., Arora, M.K. and Varshney, P.K. 2008. Multisource classification using support vector machines: An empirical comparison with decision tree and neural network classifiers. *Photogrammetric Engineering and Remote Sensing*, Vol. 74, No. 2, pp. 239-246.

Wurm, M.,Taubenböck, H., Schardt, M.,Esch, T., and Dech, S. 2011. Object-based image information fusion using multisensor earth observation data over urban areas. *International Journal of Image and Data Fusion*, Vol. 2, No. 2, pp. 121 - 147

Xu, H. and Li., P. 2010. Urban land cover classification from very high resolution imagery using spectral and invariant moment shape information. *Canadian Journal of Remote Sensing*, Vol. 36, No. 3, pp. 248-260.

Yu, B. Liu, H, Wu, J., Hu, Y. and Zhang, L. 2010. Automated derivation of urban building density information using airborne LiDAR data and object-based method. *Landscape and Urban Planning,* Vol. 98, No. 3-4, pp. 210–219.

Zhang, Y., 1999. Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. *ISPRS Journal of Photogrammetry & Remote Sensing,* Vol. 54, No.1, pp. 50–60.

Zhang, Q. and Couloigner, I. 2006. Benefit of the angular texture signature for the separation of parking lots and roads on high resolution multi-spectral imagery. *Pattern Recognition Letters,* Vol. 27, No. 9, pp. 937-946.

Zhou, W., Huang, G., Troy, A. and Cadenasso, M.L. 2009. Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: A comparison study. *Remote Sensing of Environment*, Vol. 113, No. 8, pp. 1769-1777.

Zhou, W. and Troy, A. 2008. An Object-oriented approach for analyzing and characterizing urban landscape at the parcel level. *International Journal of Remote Sensing*, Vol. 29, pp. 3119-3135.

Zhou, W., Troy, A. and Grove, M. 2008. Object-based land cover classification and change analysis in the Baltimore metropolitan area using multitemporal high resolution remote sensing data. *Sensors*, Vol. 8, No. 3, pp. 1613-1636.

Chapter 3 : OBJECT-BASED CLASSIFICATION OF URBAN AREAS USING VHR IMAGERY AND HEIGHT POINTS ANCILLARY DATA[1]

Abstract

Land cover classification of very high resolution (VHR) imagery over urban areas is an extremely challenging task. Impervious land covers such as buildings, roads, and parking lots are spectrally too similar to be separated using only the spectral information of VHR imagery. Additional information, therefore, is required for separating such land covers by the classifier. One source of additional information is the vector data, which are available in archives for many urban areas. Further, the object-based approach provides a more effective way to incorporate vector data into the classification process as the misregistration between different layers is less problematic in object-based compared to pixel-based image analysis. In this research, a hierarchical rule-based object-based classification framework was developed based on a small subset of QuickBird (QB) imagery coupled with a layer of height points called Spot Height (SH) to classify a complex urban environment. In the rule-set, different spectral, morphological, contextual, class-related, and thematic layer features were employed. To assess the general

[1] This paper has been published in Journal of *Remote Sensing* :
Salehi, B., Zhang,Y., Zhong,M., Dey, V., 2012. Object-based Classification of Urban Areas using VHR Imagery and Height Points Ancillary Data, *Remote Sensing* ,4(8): 2256-2276.

applicability of the rule-set, the same classification framework and similar one using slightly different thresholds applied to larger subsets of QB and IKONOS (IK), respectively. Results show an overall accuracy of 92% and 86% and a Kappa coefficient of 0.88 and 0.80 for the QB and IK Test image, respectively. The average producers' accuracies for impervious land cover types were also 82% and 74.5% for QB and IK.

Keywords: object-based classification; very high resolution imagery; multisource data; urban land cover; misregistration; transferability

3.1 Introduction

With the availability of VHR satellite imagery (spatial resolution ≤ 1m) since 1999, urban land cover classification using this type of data has become an emerging field of research in the remote sensing community. Because of the sub-meter spatial resolution, VHR imagery has a very high potential in more detailed and accurate mapping of urban areas [Pacifici *et al.,* 2009]. However, the high spectral variation within the same land cover type (within-class) and the low spectral variation between different land cover types (between-class) in urban areas make the classification very challenging if the classifier relies solely on spectral information of the image [Mohapatra *et al.,*2010; Myint *et al.,*2011; Blaschke *et al.,* 2011].

To differentiate impervious urban land covers such as buildings, roads, and parking and paved areas additional information should be incorporated into the classification process. Additional information could be obtained from the spatial measures extracted either from the image, in the forms of textural, morphological, and contextual measures, or from

ancillary data [Salehi *et al.,* 2012]. Over the past decade, a significant amount of research has employed spatial measures extracted from the image (i.e. texture, context, and morphology) in the classification process of VHR imagery over urban areas [Shackelford *et al.,* 2003; Binaghi *et al.,* 2003; Puissant *et al.,*2005; Carleer *et al.,* 2006; Xu *et al.,*2010; Salehi *et al.,* 2011a; Salehi *et al.,* 2011b].

The incorporation of ancillary data such as LiDAR data, digital elevation models extracted from stereo optical imagery, and vector data together with VHR imagery, which is often called multisource classification [Watanachaturaporn *et al.,* 2008], has received increasing attention in the remote sensing community in recent years. A bibliographic review of multisource data fusion for urban remote sensing applications is presented in [Gamba *et al.,* 2005]. Examples of integrating LiDAR data and VHR imagery for urban land cover mapping are provided in Watanachaturaporn *et al.,* [2008], Matikainen *et al.,* [2011] and Zhang [2010]. Despite the wide availability of vector data, few studies have benefited from these types of data in land cover classification of VHR imagery over urban areas. An example is the work carried out by [Thomas *et al.,* 2003] in which some spatial modelling techniques of vector data (road centre lines and parcel layers) were developed to deal with the problem of confusion between spectrally similar classes. Another example is [Moskal *et al.,* 2011], who employed road maps and building footprints in segmenting high resolution aerial photographs for monitoring urban tree cover.

A problematic issue in multisource image analysis is the misregistration between layers from different sources (e.g., VHR image and vector data). A precise geometric

registration of corresponding data layers is often very difficult to achieve, particularly in VHR imagery. Because of the sub-meter pixel size of VHR imagery, a pixel-by-pixel co-registration between vector data and VHR imagery is near to impossible. For this, pixel-based classification approaches do not yield promising results for multisource classification. Object-based approaches, on the other hand, facilitate the use of ancillary data [Kim *et al.,* 2010] and since they require less precise registration of data, object-based approaches are highly desirable for multisource image analysis [Zhou *et al.,*2009]. In object-based classification the basic element is a group of pixels (segments) instead of a single pixel. Consequently, it is not required that each pixel is exactly co-registered with the corresponding pixel in another layer. Indeed, if the objects of two different layers have reasonable overlap, a small shift between different data layers can be ignored during the classification process. In addition, the basic processing unit of object-based classification is an aggregated group of pixels forming a semantically meaningful object polygon which thus can effectively be integrated and analyzed with vector data [Zhang 2010].

Object-based image analysis also facilitates the incorporation of spectral and spatial information inherent in the VHR imagery into the classification process. By segmenting the image, several spectral, textural, and morphological (e.g. shape and size) features of objects can be incorporated into the classification process. These objects' features help the classifier to distinguish spectrally similar land cover types (e.g. buildings and roads). In addition, object-based classification reduces high spatial frequency noise present in

VHR images by exploiting the spectral and spatial dependency of neighbouring pixels, in the form of objects, and thus increasing the classification accuracy.

The primary objective of this research was to develop an object-based classification framework using the integration of VHR imagery and vector data such as Spot Height (SH) layer to classify an urban environment comprised of large buildings, small houses, parking lots, roads/streets, and vegetation including grass and trees. The second objective was to assess the general applicability and transferability of the framework to different areas and different VHR imagery. Finally, the third objective was to evaluate the effect of possible misregistration between the vector data and VHR images of various geometric accuracies (e.g., IK and QB) on object-based classification.

To do this, a hierarchical rule-based object-based classification framework was developed using the Cognition Network Language available in the eCognition® software package. The rule-set was developed using a small subset of QB imagery by combining different spectral, morphological (geometry and extent), contextual, and class-related features of objects, resulting from the segmentation step, together with the information of a SH layer. To assess the transferability of the developed rule-set to different areas and images, the same and similar rule-set was applied to a different and larger area of the QB and IK image, respectively. Both QB and IK image have significant and different misregistration (about 10 meters) with the SH layer. To evaluate the effect of misregistration on classification accuracy, misregistration of these images was deliberately left uncorrected.

This paper is structured as follows: Section 3.2 describes data used and the study area. The proposed methodology is presented in section 3.3. Section 3.4 and 3.5 present results and discusses, respectively. Finally, section 3.6 gives the conclusion of this study.

3.2 Datasets and Study Areas

3.2.1 Study Area

Two different parts of the city of Fredericton in New Brunswick, Canada were chosen as the study areas in this research. These include a small part of the city on the north side (Figure 3.1a) that was used for developing the rule-set and a larger part of the city on the south side that was used for testing the proposed method (Figure 3.1b). The city contains a variety of urban land cover types including vegetation areas (grass and trees), water, large buildings, small houses, parking and paved areas (with various sizes, shapes, and colors), narrow streets, and highways. Classification of such areas is challenging due to the complexity of land cover types. There are many buildings and small houses with a variety of roof colors such as white, gray and black. In residential areas, small houses and narrow streets are partially covered by trees foliage since the images were collected in mid and late summer, when the trees have reached their maximum growth. Shadows cast by tall buildings are another source of land cover complexity.

3.2.2 Datasets

Two sets of data covering the same area are used in this study. These are a) subsets of QB and IK and b) a vector layer of height points called Spot Height (SH).

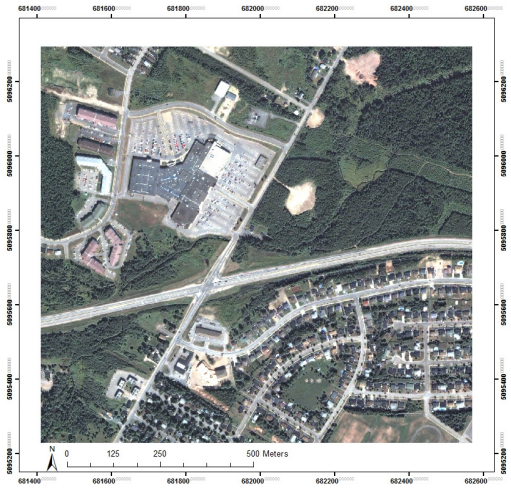
3.2.2.1 The QB and IK image

The IK and QB images were acquired on June 19 and August 31, 2002 respectively. Both QB and IK imagery possess a panchromatic band (Pan) and four multispectral (MS) bands including blue (B), green (G), red(R), and near infrared (NIR). The middle wavelengths of QB multispectral bands are 487.5, 543, 650, and 816.5 nm for B, G, R, NIR bands, respectively, while those of IK are 480.5, 550, 665, and 805 nm. The spatial resolution of pan band is 0.65 m (QB) and 0.82 m (IK) at nadir, while for MS bands, it is 2.62 m for QB and 3.28 m for IK [DigitalGlobe 2009; GeoEye 2009]. As the preprocessing step (for both VHR imagery), the four multispectral bands were fused with the panchromatic band introduced by [Zhang 2004] resulting in four pan-sharpened bands with a spatial resolution of 0.7 m and 1 m for QB and IK. Figure 3.1a shows the portion of the QB image used for developing the rule-set (QB Pilot), and Figure 3.1b shows the QB Test image.

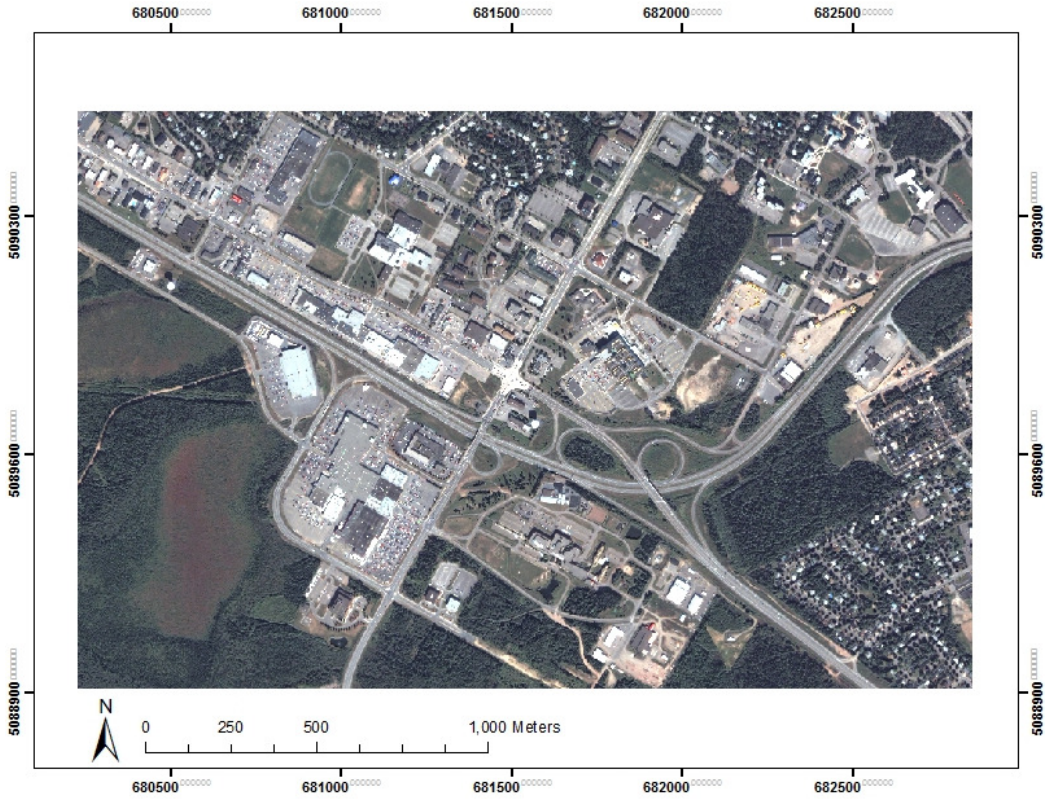
3.2.2.2 Spot Height Vector Data

The SH data was a by-product of a digital elevation model (DEM) generation project conducted by the City of Fredericton in 2002. The spot heights were collected by the City of Fredericton from 2002 stereo aerial photography in which the photo scale was approximately 1:10 000. The points were collected photogrammetrically as a series of irregularly spaced points with spacing of approximately 20 meters in both dimensions and vertical accuracy of +/- 0.40 meters at a confidence level of 90%. These points were extracted using stereo plotters in which a clear view of the ground was possible (e.g., no points in dense forest areas). In addition, the points mainly lie in non-built-up areas.

The geometrical shift (misregistration) between the SH layer and QB and IK poses a challenge when they are integrated for the subsequent image analysis (e.g., classification). This data was collected in 2002 and, unfortunately, we could not find the statistical information about the misregistration between these data layers. However, visual inspection reveals a slight misregistration between the SH layer and QB (Figure 3.2a), and a significant misregistration between the SH layer and IK. A number of points were manually collected over both QB and IK image to determine the approximate shift between these two images. A misregistration shift of approximately 10 meters between the QB and IK image was calculated. Although this amount of shift is relatively high and can be mitigated by a few ground control points, no effort was made to mitigate it as one of the primary objectives of this study is to assess the effect of misregistration on classification. It should be noted that only the locations of Spot Heights were used and no elevation information of SH layer was used in this study.



(a)



(b)

Figure 3.1 (a) Pan-sharpened QB Pilot image used for developing the rule-set. **(b)** QB Test image in true color composite.

(a) (b)

Figure 3.2 Misregistration between the SH layer and the QB (a) and IK (b) image. Asterisks represent the Spot Heights. Spot Heights were collected over non built-up areas; however, due to misregistration, some points lie on building roofs as shown in this Figure. As seen, misregistration is larger in IK than in QB.

3.3 Methodology

The proposed methodology is illustrated in Figure 3.3. The rule-based expert system starts by segmenting the Pan-sharpened image followed by a four-step hierarchical classification. First, the entire image is classified to vegetation and non-vegetation. Second, vegetation areas are further broken down to grass and trees. Non-vegetation areas are also classified to shadows and non-shadow in this step. Third, non-shadows are divided to parking lots and non-parking lots. Finally, non-parking lots are classified to buildings and roads. The remainder of this section describes the details of the flowchart in Figure 3.3.

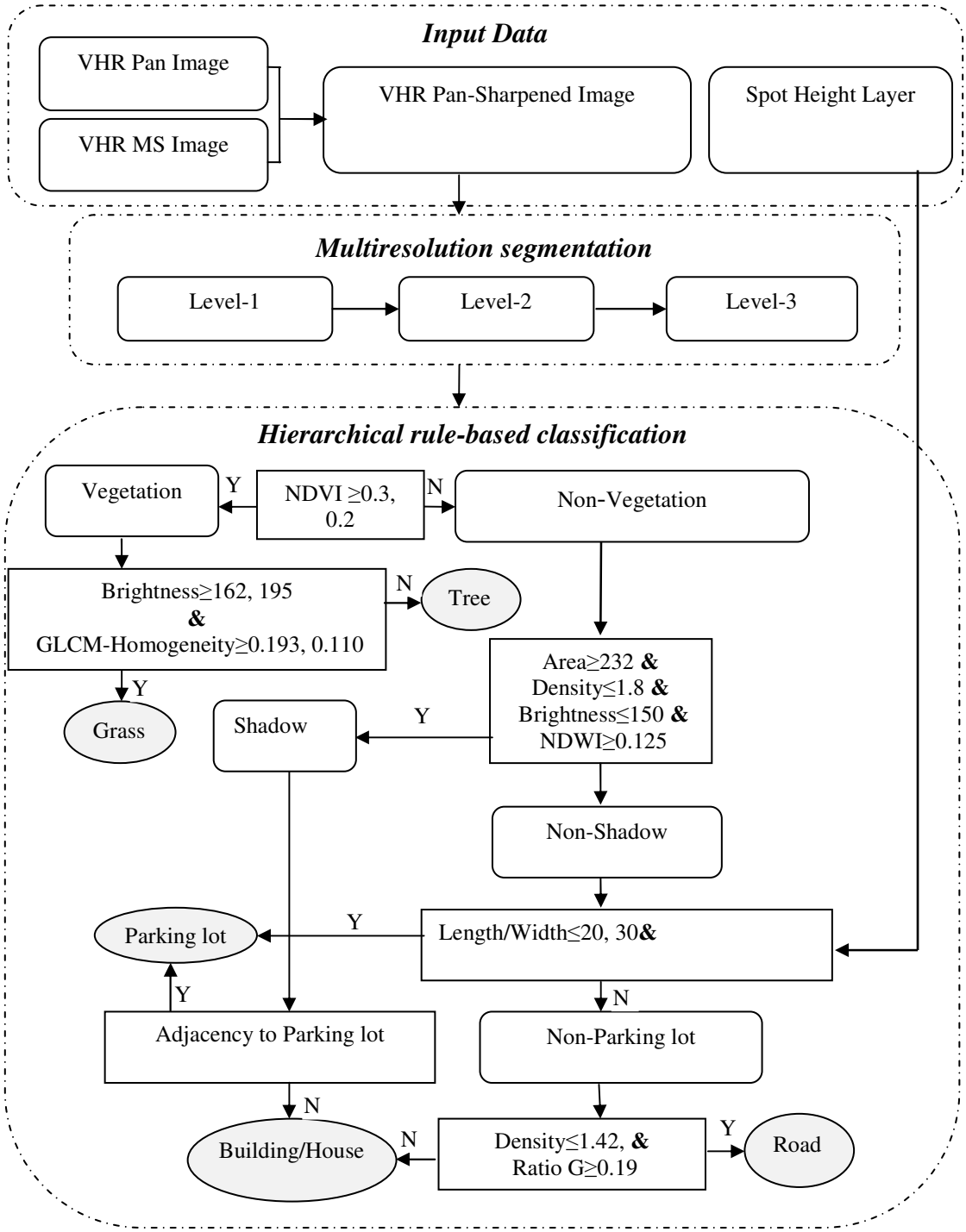


Figure 3.3 Flowchart of the developed rule-based expert system. Final classification results are shown in oval shape. The first and second thresholds for some features represent the threshold for QB and IK, respectively. When the threshold is identical for both images only one value is provided.

3.3.1 Image Segmentation

The first step in object-based image analysis is segmentation. In general, image segmentation is defined as the process of partitioning an image into separated regions based on parameters specified [Myint *et al.,* 2011]. These parameters often consider the homogeneity/heterogeneity of regions [Pal and Pal 1993]. Depending on how homogeneity (or heterogeneity) is evaluated, as well as how the pixels are aggregated (e.g., edge contour-based and region-based), there exists a large number of image segmentation techniques [Gonzalez and Woods 2005]. One of the widely used techniques is region-based, which includes both region growing and merging and splitting [Blaschke *et al.,* 2006]. The region growing method starts from seed pixels, a bottom-up approach, and regions grow until a homogeneity/heterogeneity criterion is satisfied.

In this study, multiresolution segmentation, a region-based technique [Benz *et al.,* 2004] available in eCognition which combines both region-growing and region-merging techniques [eCognitiom 2010a], was used. In multiresolution segmentation, both spectral (color) and spatial (shape) homogeneity of objects are considered. In fact, adjacent regions are merged based on a weighted homogeneity criterion of object shape and color [Baatz and Schape 2000]. Three key parameters, namely scale, shape, and compactness need to be set in multiresolution segmentation [eCognitiom 2010b]. Scale, which is considered the most crucial parameter, controls the average size of objects and the two other parameters control the homogeneity of objects [Baatz and Schape 2000]. Finding the optimal parameters for segmentation is a trial and error process, which is very time consuming and directly depends on the analyst's experience [Zhang *et al.,* 2010]. Instead

of trial and error, we used the Fuzzy-based Segmentation Parameter optimizer (FbSP optimizer) developed by [Zhang *et al.,* 2010; Tong *et al.,* 2012] to get proper parameters in different levels of segmentation. FbSP optimizer is a supervised approach for automatic estimation of the three optimal segmentation parameters (scale, shape, and compactness) using the spectral and spatial information of training objects utilized in a fuzzy interface system. It is based on the idea of discrepancy evaluation to control the merging of sub segments to reach a target segment [Tong *et al.,* 2012].

To use the FbSP optimizer, an initial segmentation is carried out by manually selecting the parameters (level 1). Normally the eCognition defaults are used for shape and compactness, and the scale parameter is set in such a way that the resulting objects are smaller than the real objects (small scale). After the first level of segmentation, a few objects (e.g. objects that represent a building) are selected as training objects (similar concept to selecting training pixels in traditional supervised classification). The information of training objects such as texture, brightness, area, and rectangular fit [Zhang *et al.,* 2010] are used to train the FbSP optimizer. After the training, the FbSP optimizer gives the optimal parameter for the second level of segmentation. Again, objects in the second level are used as training objects for calculating the parameters for the third level, and this process is iterated until the software gives objects which are close to the real objects. The full description of the FbSP optimizer can be found in [Zhang *et al.,* 2010] and [Tong *et al.,* 2012].

3.3.2 Image Classification

The second step in object-based image analysis is classification. Initially, five land covers

were defined: vegetation, shadows (excluding the shadow of trees), parking lots, roads (including wide and narrow roads, highways, and streets), and buildings (including large buildings and small houses). Vegetation was then broken down to trees and grass. Also, shadows were later assigned to either parking lots or buildings and thus the final classification map contains five land cover types including three impervious land covers, grass, and trees. A hierarchical rule-based classifier was developed to assign each object to a land cover class. The object-based approach allows the analyst to combine spectral, textural, morphological (geometry and extent), contextual, and class-related features of objects in order to assign a class membership degree (between 0 and 1) to each object based on a fuzzy membership function [Benz *et al.,* 2004 ; Walker and Blaschke 2008]. Furthermore, it has a hierarchical capability to classify the entire scene into general classes (e.g., vegetation and non-vegetation areas). These general classes are called parent classes. Then, each parent class is divided to sub classes (child class) containing more detailed land cover types (e.g., buildings and roads). This hierarchical capability allows the developer to incorporate objects in different levels of segmentation for individual levels of class hierarchy.

Table 3.1 reports image object features employed in the hierarchical rule-set for classifying the Pilot image into the aforementioned five classes. The criteria used for each class will be described in more detail in the following subsections.

Table 3.1 Object's features used in the rule-set hierarchy for different classes

Class	Segmentation level	Features
Vegetation	L1	NDVI
Shadow	L1	Brightness, Area, Density, NDWI
Grass and Trees	L2	Brightness, GLCM-Homogeneity
Parking lot	L3	Number of SH points laid within the boundary of objects, Length/width of main axis, Existence to shadow
Road/Street	L3	Density and Ratio G
Building/House	L3	Unclassified Objects, Existence to shadow

3.3.2.1 Vegetation

The first step in the proposed hierarchical expert rule-based system is to extract vegetation. Active vegetation can be identified in the near infrared spectrum due to the rise reflectance value compared to the red spectrum. Reflectivity rises sharply at 0.75 μm, the so called red-edge region [Mather 1999]. The normalized difference vegetation index (NDVI) has been widely used in the literature to separate vegetation from non vegetation areas. It is calculated by:

$$NDVI = (NIR - R)/(NIR + R)$$

Where NIR and R are the mean values of all pixels (within the boundary of each object) in band near infrared and red for a given object in each level of segmentation. Based on our experience the threshold for classifying vegetation areas was set to 0.3 in the Pilot image. Since the spectral properties of the original image (e.g., NDVI) are better preserved in smaller objects than in larger ones, the lowest level of segmentation (L1) is preferred for extracting vegetation.

3.3.2.2 Grass and Trees

Having extracted the vegetation areas, they were further classified into grass and trees. These two classes are distinguishable from their textural characteristics. The grey level co-occurrence matrix (GLCM) introduced by Haralick [1979] is the most commonly used texture measure for the urban land cover classification of VHR imagery. After visually examining different GLCM texture measures, we found that the Homogeneity measure can effectively separate grass and trees. Grass is more homogenous than trees and thus has higher Homogeneity values. This measure can be calculated for individual objects in each band. Since the spectral reflectance of vegetation is larger in NIR than in any other band, NIR was selected for the GLCM texture calculation. In addition, shadows of trees cast on grass have a texture close to that of trees. Hence, the *Brightness* values of objects were also employed in the rule-set to distinguish shaded grass from trees. The *Brightness* value is calculated as the average mean values of all bands over each object [eCognition 2010a]:

$$Brightness = (B + G + R + NIR) / 4$$

GLCM texture measure is more meaningful for objects in level 2 than level 1 and thus second level of segmentation is preferred for grass and trees classification.

3.3.2.3 Shadows

The second step in the classification hierarchy includes separating grass from trees and also extracting shadow from non-vegetation. Shadows are dark features in optical imagery, and because the segmentation of such imagery is mainly based on the spectral structure of the image [Smith and Morton *2010*], shadow influences the segmentation of

surrounding areas. Furthermore, shadow is an important factor that reduces the spectral values of the shaded objects and thus influences the land cover classification [Lu *et al.,* 2010]. For this, it should first be extracted and excluded from the subsequent segmentation and classification. Later, the area under shadow will be assigned to parking lots or buildings based on its adjacency to these classes. As mentioned, we only considered the shadow of buildings and not those from trees, since trees' shadows were included in the class of either trees or grass.

Regarding the spectral and spatial properties of shadow, two spectral (NDWI and Brightness) and two morphological features (Density and Area) of image objects were used in classifying the shadow areas.The normalized difference water index (NDWI) [Mcfeeters 1996] is a metric used for masking out black bodies (water and shadow) in VHR imagery [Chen *et al.,* 2009] and is defined as follows [Mcfeeters 1996]:

$$NDWI = (G - NIR)/(G + NIR)$$

Where G is the average mean value of all pixels (within the boundary of each object) in each level of segmentation for the band green. If only the spectral features of objects are utilized for extraction of shadow areas, some buildings, and especially small black roofed houses, are also misclassified as shadow. Thus, shadows extracted by spectral measures, must be refined to exclude spectrally similar non-shadow areas. This is feasible by employing the morphological features in the classification decision. As mentioned, *Density* and *Area* of objects were used as the morphological features to refine the shadow areas. In eCognition, the *Density* feature describes the distribution, in space, of the pixels of an object and is calculated by the Area of the object divided by the approximated

average of the major and minor axes of the ellipse fitted to the object [eCognition 2010a]. *Area* is the number of pixels forming an image object. The size of shadows in optical imagery depends primarily on the height of objects but also on the sun elevation angle [Dare 2005].In our study area (a typical North American small city), majority of buildings are two to three stories high. Therefore, shadows are normally small features compared to surrounding features such as buildings and parking lots, and thus objects in the first level of segmentation better represent shadow areas.

3.3.2.4 Parking lots

Having extracted vegetation and shadows, they are excluded from the subsequent processes in the rule-set hierarchy. Objects in the first level of segmentation do not represent the boundaries of impervious land cover types. Because of the relatively large size of parking lots, objects in the third level were chosen for the subsequent classification process. Parking lots and roads are spectrally similar because the same materials (e.g., asphalt, gravel, etc) are used in their construction. Parking lots and buildings are also similar in terms of their morphological features such as shape and extent. Therefore, extracting the parking lots solely based on the information of the image does not yield promising results. The SH layer which contains points in areas other than those with roads and buildings was utilized to help the classifier in distinguishing parking lots from roads and buildings. For classifying parking lots, the SH layer was intersected with the objects in level 3. A threshold of three points within each object was set (by visual inspection) to extract objects containing three or more height points and to classify them as parking lots. Furthermore, due to the misregistration between the VHR imagery

and SH layer, roads which are neighboring parking lots may be misclassified as parking lots. Therefore, another criterion was defined based on the *Extent* feature of objects to reduce the possible misclassification of roads as parking lots. The ratio between the length and width of the main line of each object's skeleton [eCognition 2010a], which has a high value for roads and a low value for parking lots, was considered as an *Extent* feature for parking lot classification as seen in the flowchart of the method in Figure 3.3.

3.3.2.5 Roads and Buildings

The last step in the hierarchical rule-based classifier is to classify roads and buildings. Separation of these two impervious land covers is feasible with the employment of proper object features and level of segmentation. Visual inspection of different segmentation levels confirms that roads and buildings are well represented in the third level of segmentation (L3). Moreover, the linear structure of roads, compared to the compact structure of buildings, helps the classifier to distinguish between these two classes. After visually examining several spectral features of objects including the *Brightness* and Ratios of all four bands, we found that *Ratio G*, a spectral feature, together with *Density*, a shape feature, can effectively distinguish between roads and buildings. *Ratio G* describes the amount that band G contributes to the total brightness for an object and is calculated as follows [eCognition 2010a]:

$$RatioG = G/(B + G + R + NIR)$$

Having classified roads, the remaining unclassified areas are assigned to the class of buildings. The diversity in shape and color of buildings is relatively higher than of any

other class in the scene. For this, we left buildings as the last class to be extracted in the hierarchical classification system.

3.3.2.6 Classifying Shadows

After extracting shadow areas, they should be assigned to the corresponding land cover class. Visual inspection of the image reveals that shadows cast by tall buildings belong to either parking lots or buildings. A few buildings have multi level roofs and the shadow of the top roof covers part of the lower roofs. The rest of the shadows cover part of the parking lots surrounding the buildings. Shadow areas can possibly cover the road, but since the majority of shadows in the image are cast by large and tall buildings, which are normally encompassed by parking lots, most of the shadow areas belong to parking lots. Therefore, in the rule-set shadows were assigned to either buildings or parking lots. The assignment of shadows was carried out through the employment of class-related features. The class-related feature used in this study is the *existence of neighbor objects*. In other words, if the shadow is adjacent to parking lot object(s), it is assigned to parking lots; otherwise it is assigned to the class of buildings. It should be noted that this rule is hardly transferable to areas with different urban structure. For instance, shadows may belong to road as well. However, in our study area, shadow does not cover a large area and missing some shadow does not have a significant effect on the classification accuracy of the entire image. Moreover, further development of rules for shadow will lead to a more complex rule-set and thus affect the transferability of the entire rule-set to other areas.

3.4 Results

3.4.1 Segmentation Results

The segmentation results of the three images, using the FbSP optimizer, are reported in Table 3.2. The visual results of a portion of the QB and IK Test image are also depicted in Figure 3.4. The three segmentation parameters (i.e., scale, shape, and compactness) are different for QB and IK image. This is mainly because of the slight difference between the spatial resolution, the wavelength coverage of each multispectral band, and the acquisition date of QB and IK image (see section 3.2.2).

Table 3.2 Multiresolution segmentation parameters for the three images

	Level	**Scale**	**Shape**	**Compactness**	**No of Objects**
QB-Pilot image	1	30.00	0.10	0.50	18204
	2	77.83	0.64	0.81	2190
	3	131.33	0.50	0.81	912
QB-Test image	1	30.00	0.10	0.50	64481
	2	77.83	0.64	0.81	7634
	3	131.33	0.50	0.81	2793
IK-Test image	1	40.0	0.10	0.50	69102
	2	82.0	0.57	0.80	10810
	3	142.0	0.59	0.80	2908

Objects in the first level of segmentation are not meaningful for any of the land covers. However, since the original spectral properties of the image are better preserved in lower levels than in higher levels of segmentation, the lower levels (i.e., L1) preferred for classifying land covers in which the spectral features are employed (e.g., shadow and vegetation).

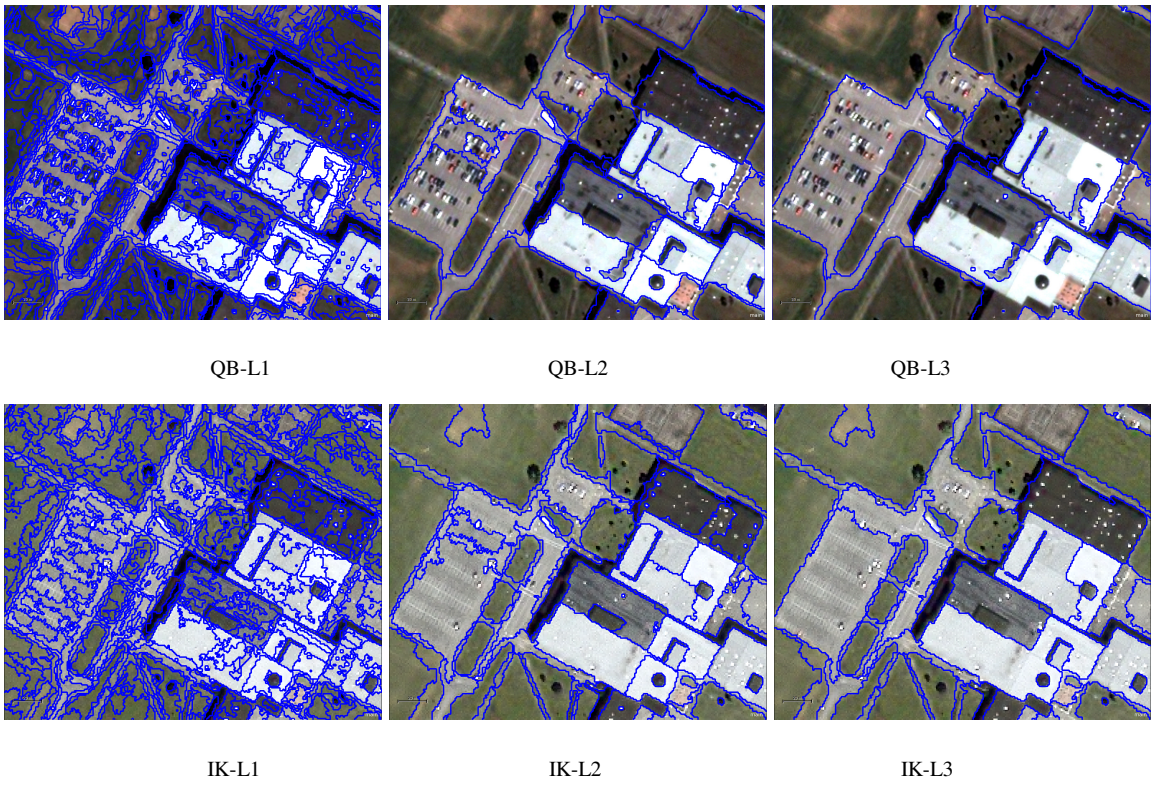


Figure 3.4 Multiresolution segmentation results of level 1 (L1), level 2 (L2), and level 3(L3) for a sample area of QB (top) and IK (bottom) Test images.

For the classification of impervious land covers, larger objects (level 3) proved to be more effective than smaller objects. There are two main reasons for this First, morphological features such as size and shape play a more important role than spectral features for the classification of such land covers. Furthermore, shape and size of impervious land covers are better modeled in higher levels of segmentation where objects are more meaningful. Second, the error of misregistration between different data layers (e.g., vector data and the VHR image) has less impact on larger objects than for smaller objects. In fact, if the object of interest has a reasonable overlap in the two layers, which is more likely for larger objects, the misregistration can be ignored.

3.4.2 Classification Results

In order to evaluate the quality of land cover classification, accuracy assessment is often required. The conventional pixel-based *error matrix* [Foody 2005; Congalton and Green 2009] was used for accuracy assessment of the results. Different elements of the error matrix including overall accuracy, kappa coefficient, user's and producer's accuracy were calculated for both QB and IK Test image. Kappa coefficient is calculated based on the difference between the actual agreement in the error matrix (i.e. agreement between the reference data and classification result as indicated by the major diagonal) and the chance agreement that is indicated by the row and column totals (i.e. marginals) [Congalton and Green 2009]. For each dataset, the Z-test was also performed to test the significance of the error matrix. The Z-test checks whether the classification is meaningful and significantly better than a random classification [Congalton and Green 2009]. For each Test image, the result of object-based land cover classification was exported in raster and was compared with reference data. The reference dataset was generated through a precise manual digitizing of many samples (60% of the entire image) of five land cover types on both the QB Pilot, and the QB and IK Test image.

3.4.2.1 Classification of the Pilot image

As mentioned earlier, the QB Pilot image was used to develop the hierarchical expert rule-based classification system. Figure 3.5 shows the classification results of the Pilot image. The confusion matrix and its measures including overall accuracy, producers' accuracies, users' accuracies, and the kappa coefficient are reported in Table 3.3. For the

purpose of transferability and the general applicability of the rule-set to other areas and images, few spectral and spatial features of the objects were utilized in the classification of the Pilot image. The fewer number of objects' features leads to a less complex rule-set, and consequently it better guarantees the transferability of the same rule-set to other images. Notwithstanding the low complexity of the developed rule-set, the classification result of the Pilot image is very promising. An overall accuracy of 95% and a Kappa coefficient of 0.92 were achieved for this image (Table 3.3). Trees, grass, and roads were classified with producers' accuracies of more than 92%. Although there are a few misclassifications between buildings and parking lots, the majority of these two classes were correctly classified. The accuracies for buildings and parking lots were 84% and 89%, respectively.

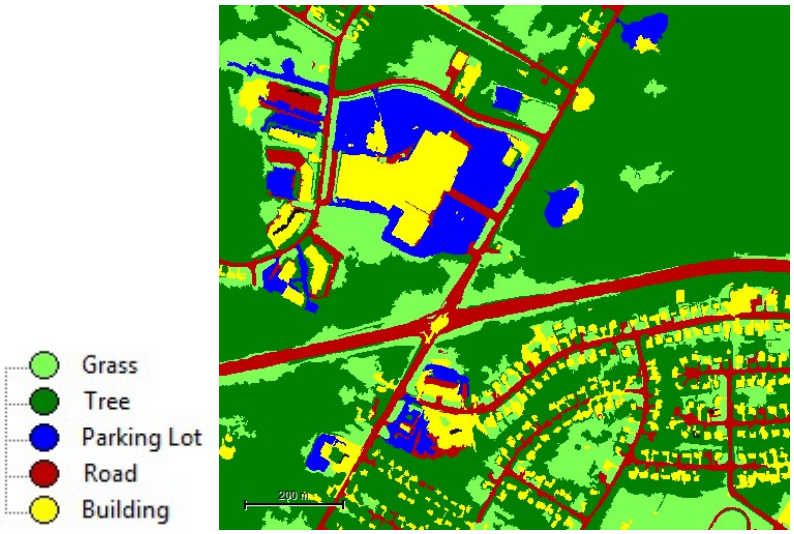


Figure 3.5 Hierarchical object-based classification results of the QB Pilot image

Table 3.3 Classification results for the QB-Pilot image

		Reference Data (No. of Pixels)					
	Class name	Tree	Grass	Building	Road	Parking lot	User's Acc.(%)
Classification Results	Tree	778395	10887	12427	6600	1313	92.9
	Grass	9587	189105	1634	1837	1327	96.1
	Building	184	1065	141261	1530	11366	90.9
	Road	225	2107	8254	152881	3362	91.6
	Parking lot	102	567	4274	2534	140175	94.9
	Prod's Acc.(%)	98.7	92.8	84.2	92.4	89.0	
	Overall Accuracy: 94.53% Kappa Coefficient : 0.92 Z Statistic: 3251						

3.4.2.2 Classification of the Test images

The rule-set that was used for the segmentation and classification of the QB Test image was identical to the one used for the Pilot image. The same workflow and object features were also used for the IK Test image. However, due to the difference between QB and IK image in terms of spatial resolution, wavelength range of multispectral bands, the date of acquisition, and also the segmentation results, different thresholds for object's features were used in classifying the IK Test image. The confusion matrices and their measures are reported in Tables 3.4 and 3.5. The results are also displayed in Figure 3.6.

At the 99.9 % confidence level, the critical value for Z statistic would be 3.3. Therefore, if the Z value of the classification results is greater than 3.3, the results are significant and better than random. The Z values for QB and IK are far greater than the critical value (5434 for QB and 2731 for IK) showing that the classification is significantly better than a random classification. The kappa coefficient of 0.88 and 0.80 for the QB and IK Test image represent a strong agreement between reference data and classification results. The overall classification accuracy of QB is 91.6%, confirming the high potential of

transferability of the developed rule-set to different areas of the same image. For IK, however, a lower overall accuracy of 85.6% was achieved. The comparison of error matrices between IK and QB shows that the big difference is between the accuracies of the impervious land covers, especially parking lots. This is mainly because of the large misregistration between the SH layer and the IK image (Figure 3.2), since this layer is the key feature in classifying the impervious land covers. Nevertheless, this level of accuracy for IK shows the efficiency of the proposed method in classifying urban areas of different VHR imagery.

The highest producer's accuracy for both images was achieved for the class of trees and grass with 98.8% and 93.9%for QB and 95.3% and 91.2% for IK. For these two classes the user's accuracy is also highest among all classes. These indicate the very high potential of object-based classification of VHR imagery for extracting vegetation areas using the well known index of NDVI and segmentation in lower levels (L1 and L2). The selection of the threshold values for NDVI is critical in classifying vegetation areas. In this study, the thresholds of 0.30 and 0.20 were selected for QB and IK, respectively. The difference between NDVI values of QB and IK is mainly because of the difference between the acquisition dates and the objects' sizes and shapes (Table 3.2) of the two images.

The average producer's accuracy for impervious land covers for QB is 82%, while for IK it is 74.5%. Among the three impervious land covers, roads achieved the highest producer's accuracy in both images (90% for QB and 87% for IK). Roads are elongated features and they are distinguishable from buildings and parking lots using the

morphological features of objects. Additionally the within-class spectral heterogeneity of the class of road is far less than those of buildings and parking lots. The comparison of the original image (Figure 3.1b) with the classification results (Figure 3.6) reveals that all major roads and the majority of minor roads, with a few mis-classifications of road intersections, were properly classified in both images. Nevertheless, the relatively low user's accuracy of roads (73.4% for QB and 63.5% for IK) indicates that there is overestimation of roads in both images. Some parking lots and buildings have been classified as roads. This is mainly because of the spectral similarity of roads to the other two impervious land covers.

Parking lots are typically spectrally similar to roads and morphologically similar to building roofs. Therefore, the separation of parking lots from the two other classes is quite problematic. In this study, however, the key feature for extracting parking lots was the SH layer. Although there is a slight misregistration between QB and the SH layer (Figure 3.2a), utilizing the objects instead of the individual pixels in the classification process effectively mitigates this negative effect of multisource classification. The producer's accuracy of parking lots is 77% in QB. For IK Test image, however, the larger misregistration between the image and the SH layer, together with the smaller objects in level 3, leads to a lower producer's accuracy of 64% for parking lots. The FbSP optimizer resulted in a larger number of objects in level 3 for IK than for QB (Table 3.2). This means that objects in level 3 are smaller in IK compared to QB. Consequently, the possibility of overlapping three spot heights with an object (the key criteria that is used for extracting parking lots) is greater for the QB than for the IK Test image. As a result,

more parking lots are missed during classification in IK than in QB Test image leading to lower producer's accuracy. In addition, the relatively high user's accuracy of parking lots in both images (93.3% for QB and 82.6% for IK) confirms that few roads' and buildings' pixels have been misclassified as parking lots.

The last step in rule-set hierarchy is the classification of buildings. Almost all large buildings and most of the small houses in the lower left and upper part of the images were classified correctly in QB but not in the IK. The producer's accuracy for buildings in IK is 72.6%, while for QB it is 79.4%. This difference can also be interpreted based on the larger misregistration and smaller objects in IK compared to QB. The user's accuracy of buildings is relatively high for both images indicating that the error of commission is low for buildings. In other words, few roads and parking lots have been classified as buildings.

Table 3.4 Classification results for the QB Test image

		Reference Data (No. of Pixels)					
	Class name	Tree	Grass	Building	Road	Parking lot	User's Acc.(%)
Classification Results	Tree	2087435	56242	16343	14856	2441	95.87
	Grass	20517	1177125	3552	14956	1310	96.69
	Building	1991	2355	497277	15799	70516	84.58
	Road	3407	15531	83770	499652	78238	73.41
	Parking lot	251	2775	25530	7815	503632	93.26
	Prod's Acc.(%)	98.76	93.87	79.38	90.34	76.76	
	Overall Accuracy: 91.58% Kappa Coefficient : 0.88 Z Statistic: 5434						

Table 3.5 Classification results for the IK Test image

	Class name	Reference Data (No. of Pixels)					
		Tree	Grass	Building	Road	Parking lot	User's Acc.(%)
Classification Results	Tree	930657	30395	3599	7919	1705	95.5
	Grass	35590	400240	8587	3275	1967	89.0
	Building	3374	697	243902	6966	51631	79.6
	Road	6701	6984	51041	246458	76816	63.5
	Parking lot	262	644	28974	19587	235271	82.6
	Prod's Acc.(%)	95.3	91.2	72.6	86.7	64.1	
	Overall Accuracy: 85.6% Kappa Coefficient : 0.80 Z Statistic: 2731						

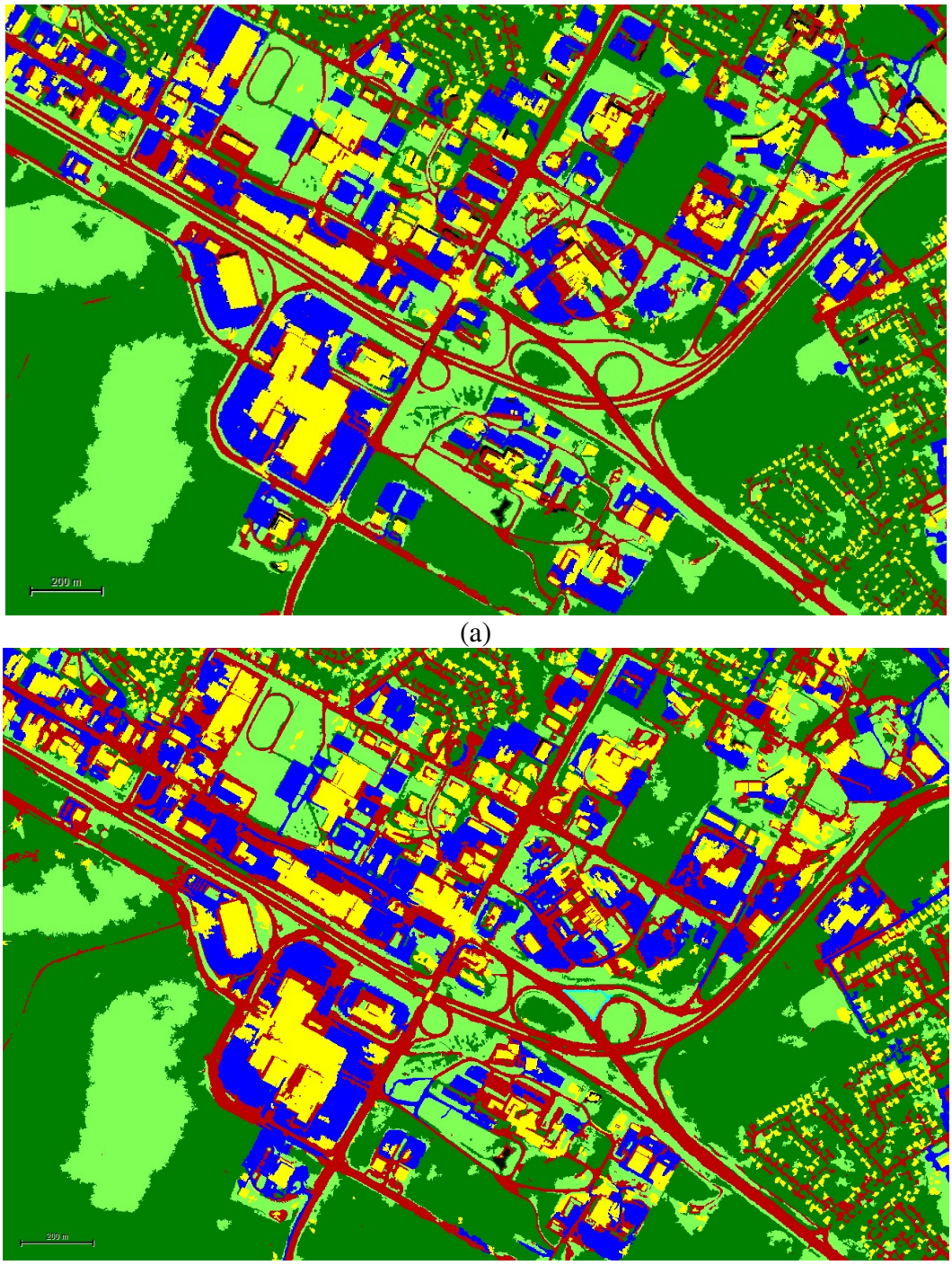


Figure 3.6 Hierarchical object-based classification results of **(a)** QB and **(b)** IK Test images.

3.5 Discussion

As mentioned, the segmentation parameters and classification rule-set of the QB Pilot image is identical to those used for QB Test image. A small discrepancy of 3% and 0.04 of overall accuracy and kappa coefficient, respectively, between the QB Pilot and Test images demonstrates the great potential of the transferability of the rule-set to other areas of the same VHR imagery. This suggests that for a large dataset, the rule-set needs to be developed using a small portion of the image and then can be applied directly to the entire dataset.

For IK Test image, however, segmentation is conducted independently of QB Pilot image. Indeed, FbSP optimizer needs to be trained for each VHR image separately. Nonetheless, the segmentation parameters of a portion of the VHR image are identical to those of the entire VHR image. For classification, the rule-set developed using the QB Pilot image is very similar to the one used for IK Test image. In other words, all the rules and features are the same but some thresholds (see figure 3) are slightly different. The discrepancy of overall accuracy and kappa coefficient between QB Pilot image and IK Test image is about 9% and 0.12, respectively. These discrepancies are not surprising as the misregistration between SH layer and IK image is significantly larger (about 10 meters) than that between SH layer and QB image (see figure 2). Checking the confusion metrics of QB Pilot and IK test images shows that the largest discrepancy is between the producers' accuracies for parking lots in two matrices (about 25%). In addition, SH layer plays a critical role in classifying parking lots and thus larger misregistration between SH

layer and the VHR image leads to worse classification result for parking lots than for any other classes.

Regardless of the misregistration effect, morphological features (e.g. density, area, etc.) prove to be more effective for standardization (transferability) of the rule-set. Most of the morphological features in the developed rule-set use the same threshold in QB and IK images. Therefore, for the general applicability of the rule-set to other images, more attention should be directed to the use of morphological features rather than spectral and/or textural features, especially for classifying impervious land cover types. However, spectral and textural features play an important role in classifying vegetation areas.

This study demonstrates the usefulness of ancillary data in conjunction with object-based image analysis for urban land cover classification of VHR imagery. The ancillary data that was used is a Spot Height data layer, which was employed for separating parking lots from buildings and roads. This layer, however, may not be available for many urban areas. Alternatively, the SH layer can be created from a relatively inaccurate digital elevation/surface model (DEM/DSM)(e.g. from stereo satellite imagery) of an urban area since the proposed method only uses the locations of Spot Height points in the rule-set but not the elevation of points. Therefore, an accurate DEM/DSM of the area which is often acquired by LiDAR data, and thus is expensive, is not required. Moreover, most VHR satellites offer stereo imagery from which DEM/DSM information can be extracted. This is a topic for future research.

3.6 Conclusion

A multisource object-based rule-based hierarchical classification approach was developed

to classify a complex urban environment. The multisource data include VHR imagery and Spot Height vector data. The rule-based classification method was first developed using a small subset of the QB imagery and then it was tested on larger areas of QB and IK imagery to assess the transferability of the rule-set to classify different areas and different images.

The method offers a practical, fast, and easy to use (within eCognition) framework for classifying VHR imagery of small urban areas. Overall accuracies of about 92% and 86 % for QB and IK are very promising, since distinguishing spectrally (buildings, roads, and parking lots) and spatially (buildings and parking lots) similar classes in urban areas is very challenging. Further, the method shows that it is applicable to different areas of the same image and different VHR images with no change in the rule-set (for the same VHR image) or slight changes in the thresholds (for different VHR images). The method also demonstrates that the impact of possible misregistration between different datasets (which is inevitable in multisource classification) on classification can be mitigated through employing object-based classification, as the basic processing unit is the object rather than the individual pixel.

The proposed framework, though it is not applicable to all urban areas, provides guidelines on the types of features (e.g. texture, shape, size, brightness) and ranges of thresholds which are suitable for classifying specific land cover types. However, ancillary data used in this study (Spot Height) may not be available for many urban areas. Utilizing more available ancillary data such as DEM/DSM is especially desirable and will be the focus of our future research.

Acknowledgments

The authors would like to thank the City of Fredericton for providing the data used in this work. This research is supported by the NSERC (Natural Science and Engineering Research Council of Canada) Strategic Project Grants and NSERC Discovery Grants Programs.

REFERENCES

Baatz, M. ;Schape, A. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T. And Griesbner, G. (Editors), *Angewandte Geographische Informations-Verarbeitung XII*, Wichmann Verlag, Karlsruhe. **2000**, pp. 12–23.

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

Binaghi, E.; Gallo, I.; Pepe, M. A cognitive pyramid for contextual classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*. **2003**, 41(12), 2906-2922.

Blaschke, T.; Burnett, C. ; Pekkarinen, A. Image segmentation methods for object based analysis and classification. In *Remote Sensing Image Analysis: Including the Spatial Domain,* Jong, S.M. de and Meer F. D. van der (eds.) (Ed.). **2006**, pp. 211-236 (Dordrecht: Springer-Verlag).

Blaschke, T. Hay, G.J., Weng, Q., and Resch, B. Collective sensing: Integrating geospatial technologies to understand urban systems-An overview. *Remote Sensing*. **2011**, 3, 1743-1776.

Carleer, A.P.; Wolff, E. Urban land cover multi-level region-based classification of VHR data by selecting relevant features. *International Journal of Remote Sensing*, **2006**, 27(6), 1035-1051.

Chen, Y.; Su, W.; Li, J.; Sun, Z. Hierarchical object-oriented classification using very high resolution imagery and LIDAR data over urban areas. *Advances in Space Research*. **2009**, 43(7), 1101-1110.

Congalton, R.G.; Green, K. *Assessing the accuracy of remotely sensed data: principles and practices, 2nd Ed.* Boca Raton, Florida: Taylor & Francis Group, **2009.**

Dare, P.M. Shadow analysis in high-resolution satellite imagery of urban areas. *Photogrammetric Engineering and Remote Sensing*. **2005**, 71(2), 169-177.

DigitalGlobe. *DigitalGlobe core imagery products guide* [online], **2009**. Available from: http://www.digitalglobe.com/index.php/6/DigitalGlobe+Products [Accessed 07 May 2012].

eCognition . *eCognition Developer (8.64.0) Reference Book*. Trimble Germany GmbH, Trappentreustr. 1, D-80339 München, Germany, **2010a.**

eCognition. *eCognition Developer (8.64.0) User Guide.* Trimble Germany GmbH, Trappentreustr. 1, D-80339 München, Germany, **2010b.**

Foody, G.M. Status of land-cover classification accuracy assessment, *Remote Sensing of Environment* **2002**, 80(1), 185–201.

Gamba, P.; Dell'Acqua, F.; Dasarathy, B. Urban remote sensing using multiple data sets: past, present, and future. *Information Fusion* **2005**, 6, 319-326.

GeoEye. *GeoEye product guide v1.0.1* [online], **2009**. Available from: http://www.geoeye.com/CorpSite/assets/docs/brochures/GeoEye_Product_Guide.pdf [Accessed 07 May 2012].

Gonzalez, R.C.; Woods, R. E.. *Digital Image Processing*, 2nd ed., **2002** (US- New Jersey, Prentice Hall).

Haralick, R.M. Statistical and structural approaches to texture. *In Proceedings of IEEE* **1979**, 67, pp. 786-804.

Kim, M.; Madden, M.; Xu, B. GEOBIA vegetation mapping in general smoky mountains national park with spectral and non-spectral ancillary information. *Photogrammetric Engineering and Remote Sensing* **2010** , 76(2), 137-148.

Lu, D.; Hetrick, S.; Moran, E. Land cover classification in a complex urban-rural landscape with Quickbird imagery. *Photogrammetric Engineering and Remote Sensing* **2010***,* 76(10), 1159-1168.

Mather, P. M. *Computer Processing of Remotely-Sensed Images: An Introduction*, 2nd Ed. John Wiley and Sons, **1999**.

Matikainen, L.; Karila, K. Segment-based land cover mapping of a suburban area—comparison of high-resolution remotely sensed datasets using classification trees and test field points. *Remote Sensing*. **2011**, 3, 1777-1804.

Mcfeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing* **1996**, 17(7), 1425-1432.

Mohapatra, R.P.; Wu, C. High resolution impervious surface estimation: an integration of Ikonos and Landsat-7 ETM+ imagery. *Photogrammetric Engineering and Remote Sensing* **2010**, 76(12), 1329-1341.

Moskal, M., Styers, D.M., Halabisky, M. Monitoring urban tree cover using object-based image analysis and public domain remotely sensed data. *Remote Sensing*. **2011**, 3, 2243-2262.

Myint, S.W.; Gober, P.; Brazel, A.; Clark, S. G. ;Weng, Q. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment* **2011***,* 115(5), 1145-1161.

Pacifici, F.; Chini, M.; Emery, W.J. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment* **2009***,* 113 (6), 1276-1292.

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

Puissant, A.; Hirsch, J.; Weber, C.The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. *International Journal of Remote Sensing* **2005***,* 26(4), 733-745.

Salehi, B.; Zhang, Y. ; Zhong, M.; Dey, V. A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery. *International Journal of GeoInformatics,* **2012**, 8 (3) pp. 35-51.

Salehi, B.; Zhang,Y.; Zhong, M. Combination of object-based and pixel-based image analysis for the classification of VHR imagery over urban areas. In the Proc. of *ASPRS 2011 Annual Conference*, May 1-5 **2011a**, Milwaukee, Wisconsin, USA.

Salehi, B.; Zhang,Y.; Zhong, M. Object-based land cover classification of urban areas using VHR imagery and photogrammetrically-derived DSM. In the Proc. of *ASPRS 2011 Annual Conference*, May 1-5 **2011b**, Milwaukee, Wisconsin, USA.

Shackelford, A.K.; Davis, C.H. A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. *IEEE Transactions on Geoscience and Remote Sensing* **2003**, 41(9 Part1), 1920-1932.

Smith, G.M.; Morton, R. W. Real world objects in GEOBIA through the exploitation of existing digital cartography and image segmentation. *Photogrammetric Engineering and Remote Sensing* **2010**, 76(2), 163-170.

Thomas, N.; Hendrix, C. ; Congalton, R.G. A comparison of urban mapping methods using high-resolution digital imagery. *Photogrammetric Engineering and Remote Sensing* **2003**, 69(9), 963-972.

Tong, H.; Maxwell, T.; Zhang, Y.; Vivek, D. A supervised and fuzzy-based approach to determine optimal multiresolution image segmentation parameters. Accepted for publication, *Photogrammetric Engineering and Remote Sensing,* February **2012**.

Walker, J. S.; Blaschke, T. Object-based land cover classification for the Phoenix metropolitan area: optimization vs. transportability. *International Journal of Remote Sensing* **2008**, 29 (7), 2021-2040.

Watanachaturaporn, P.; Arora, M.K.; Varshney, P.K. Multisource classification using support vector machines: An empirical comparison with decision tree and neural network classifiers*. Photogrammetric Engineering and Remote Sensing* **2008***,* 74 (2), 239-246.

Xu, H.; Li., P. Urban land cover classification from very high resolution imagery using spectral and invariant moment shape information. *Canadian Journal of Remote Sensing* **2010**, 36(3), 248-260.

Zhang, Y. Standard and Colour Enhanced Pansharpening of Quickbird Imagery - Sydney, Australia. *Photogrammetric Engineering & Remote Sensing* **2004** , 70(6). The Front Cover.

Zhang, J. Multi-source remote sensing data fusion: status and trends. *International journal of Image and Data Fusion*, **2010**, 1 (1), 5-24.

Zhou, W.; Huang, G.; Troy, A.; Cadenasso, M.L. Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: A comparison study. *Remote Sensing of Environment* **2009***,* 113 (8), 1769-1777.

Zhang, Y.; Maxwell, T.; Tong, H.; Dey, V. Developement of supervised software tool for automated determination of optimal segmentation parameters for ecognition. *ISPRS TC VII symposium-100 Years ISPRS*, Vienna, Austria, July 5-7, **2010**.

Chapter 4 : A COMBINED OBJECT- AND PIXEL-BASED IMAGE ANALYSIS FRAMEWORK FOR URBAN LAND COVER CLASSIFICATION OF VHR IMAGERY[1]

Abstract

Pixel-based classification approaches do not yield promising results when applied to very high spatial resolution (VHR) imagery of urban areas. This is because of the high spectral variations of pixels within the same class and the high spectral similarities between different classes (e.g., buildings and transportation areas). Object-based classification mitigates these problems by segmenting the image into groups of spectrally/spatially similar pixels called objects followed by the classification of the objects. However, the major problem associated with object-based approaches is the lack of automation and its dependency to different images and applications. In this paper a combined pixel-based and object-based method exploiting the advantages of both approaches is proposed. The method starts with segmenting the image resulting in several spectral, textural, and morphological features of segments. To overcome the "curse of dimensionality", a wavelet-based feature extraction is proposed to reduce the number of features. The wavelet-based method is automatic and fast and can preserve local variations in objects'

[1] This paper has been submitted to *Photogrammetric Engineering and Remote Sensing (PE&RS):*
Salehi, B., Zhang,Y., Zhong,M., 2012. A combined object- and pixel-based image analysis framework for urban land cover classification of VHR imagery, *Photogrammetric Engineering and Remote Sensing* (under review).

spectral/spatial signatures. Finally, the extracted features together with the original bands of the image are classified using the conventional pixel-based Maximum Likelihood classification. The proposed method was tested on the WorldView-2 (WV-2), Quickbird (QB), and IKONOS (IK) images of the same urban area for comparison purposes. Results show up to 17%, 10%, and 11% improvement in kappa coefficients compared to the case in which only the original bands of the image are used for WV-2, QB, and IK, respectively. Furthermore, objects' spectral features contribute more to increasing classification accuracy than spatial features.

4.1 Introduction

Urban land cover classification using satellite imagery has gained increasing attention in the remote sensing community since the launch of the first very high spatial resolution (VHR) satellite (i.e., IKONOS) in 1999. Unfortunately the conventional spectral-based classifiers do not offer promising results when applied to VHR imagery. The reason is that these classifiers mainly rely on the spectral information of individual pixels within the image. However, in an urban environment, the spectral heterogeneity of pixels within the same land cover type and the spectral similarity between different land cover types (e.g., buildings, roads, and parking lots) are too high making the spectral information insufficient for classification [Lu *et al.,* 2010; Huang *at al.,* 2011].

Object-based image analysis takes into consideration the spectral and spatial (e.g., texture, shape, and size) correlation of neighboring pixels by partitioning the image into groups of spectrally/spatially similar pixels called objects. Thus, not only the spectral

properties of objects but also the spatial properties such as textural, contextual, and morphological features of objects can be incorporated into the classification process. However, a major problem associated with object-based classification is its lack of automation. Object-based classification approaches are mainly rule-based relying on the analyst's experience of a particular image and application, preventing the general applicability and transferability of the approach to different areas and applications.

One way to tackle the aforementioned problems of pixel-based and object-based approaches is using a method which combines the advantages of both approaches. The method takes the spectral and spatial features of objects resulting from segmenting the image together with the original bands and classifies them using conventional pixel-based classifiers [Salehi *et al.,* 2011]. By segmenting the image, several spectral (e.g., mean brightness, ratios of bands, NDVI, etc) and spatial features such as texture (e.g., grey level co-occurrence matrices) and morphology (e.g., geometry and extent) of objects can be extracted and used in classification. In fact, we are dealing with tens of spectral and spatial features for each object which can be stacked to the original bands of the image and utilized in classification. Moreover, some segmentation methods allow multilevel segmentation resulting in the same number of features for each level [Benz *et al.,* 2004].

However, classifying this high dimensional data (spectral and spatial features) will lead to less accurate classification because of the Hugh phenomenon [Landgrebe, 2003] which was initially introduced for hyperspectral image analysis. Therefore, extracting (or selecting) proper features is a necessary step prior to the classification of such high dimensional spectral-spatial features of objects. Several feature extraction and feature

selection methods have been introduced mainly for the analysis of hyperspectral imagery. Among them are discriminant analysis feature extraction (DAFE) [Fakunaga, 1990], decision boundary feature extraction (DBFE) [Lee and Landgrebe, 1993], and multiple discriminant analysis (MDA) [Duda *et al.,* 2001]. These algorithms are all supervised (need to be trained) and are well-suited for classes with Gaussian-like distribution. However, for data with different characteristics (i.e., spectral, textural, and morphological features) the Gaussian-like assumption is hardly met. An alternative method is the well known Principal Component Analysis (PCA). PCA computes orthogonal projections that maximize the amount of data variance, and yield a new set of un-correlated image bands called components. However, for classification, these components are not always appropriate, since it is not always the case that maximum of data variance is caused by the difference between classes [Chan *et al.,* 2009]. Furthermore, PCA transformation is time consuming because of its global nature [Kaewpijit *et al.,* 2002].

In this paper, a wavelet-based feature extraction (dimensionality reduction) method is presented. The principle of this method is to apply a discrete wavelet transform to each set of spectral, textural, and morphological features of the segmented image. The wavelet is applied to the spectral/spatial signature of each individual object resulting in a new signature with significantly fewer features. The extracted features are then employed in conventional pixel-based classification. The paper also demonstrates which set of features (spectral, textural, and morphological) and which level of segmentation contributes more to increasing the classification accuracy. The method was tested on

images from different VHR satellites, WorldView-2 (WV-2), Quickbird (QB), and IKONOS (IK), acquired over the same urban areas.

The remainder of this paper is organized as follows. Section 2 presents a description of study area and VHR images used. In section 3 the methods are described. This section includes a detailed description of the segmentation of the image, spectral and spatial features extracted from segmentation, wavelet-based feature extraction, and finally the classification. Section 4 presents results and discussion. Both wavelet-based feature extraction and the corresponding classification results for different sets of data are presented and discussed. Finally, section 5 gives the conclusion of this work.

4.2 Study Area and Datasets

The study area is a part of the City of Fredericton, a typical small city in North America and the capital of the Province of New Brunswick in Canada. Five major land cover classes were defined in this study including buildings (large buildings and small houses), roads (highways, roads and streets), parking lots, grass, and trees. Classification of this area is very challenging, due to spectral and spatial complexity of land cover types. There are buildings with various roof colors and sizes. Some small houses and minor roads are partly covered by foliage. The class of parking lots is also diverse in terms of color, texture, and size (e.g., parking lots with and without cars). For comparison purposes, the images of three different VHR sensors, IKONOS, QuickBird, and WorldView-2, acquired from the same area but in different dates, were used.

The IKONOS (IK) and QuickBird (QB) images were acquired on June 19 and August 31, 2002, respectively. Both QB and IK images possess a panchromatic band (Pan) and four multispectral (MS) bands including blue (B), green (G), red(R), and near infrared (NIR). The middle wavelengths of QB multispectral bands are 487.5, 543, 650, and 816.5 nm for B, G, R, NIR bands, respectively, while those of IK are 480.5, 550, 665, and 805 nm. The spatial resolution of Pan is 0.65 m (QB) and 0.82 m (IK) at nadir, while for MS bands, it is 2.62 m for QB and 3.28 m for IK [DigitalGlobe, 2009; GeoEye, 2009].

WorldView2 (WV-2), launched October 2009, is the first VHR 8-band multispectral satellite which collects imagery with the highest spatial resolutions (together with GeoEye) among all commercially-available VHR satellite images. A subset of a geometrically corrected WV-2 image, acquired on July 20, 2011, was used in this study. WV-2 imagery comprises a panchromatic (Pan) band and eight multispectral (MS) bands. The spatial resolution of the Pan and MS bands are 0.46 m and 1.84 m at nadir, respectively. However, the image used in this study has the resolutions of 0.5 m and 2 m in the Pan and MS modes. The MS contains the four conventional bands (i.e., B, G, R, NIR1) as well as four newly added bands, which are coastal (C), yellow (Y), red edge (RE), and near infrared-2 (NIR2). The centre wavelengths of eight MS bands are at 425, 480, 545, 605, 660, 725, 835, and 950 nm for C, B, G, Y, R, RE, NIR1, and NIR2, respectively [DigitalGlobe, 2009].

In order to exploit the full spectral and spatial potential of images, the Pan band was fused with MS bands, for each image, resulting in Pan-sharpened images of IK, QB, and

WV-2. The fusion was carried out using the Pan-sharpening algorithm developed by Zhang (2004). Figure 4.1 shows the pan-sharpened WV-2 image of the study area.

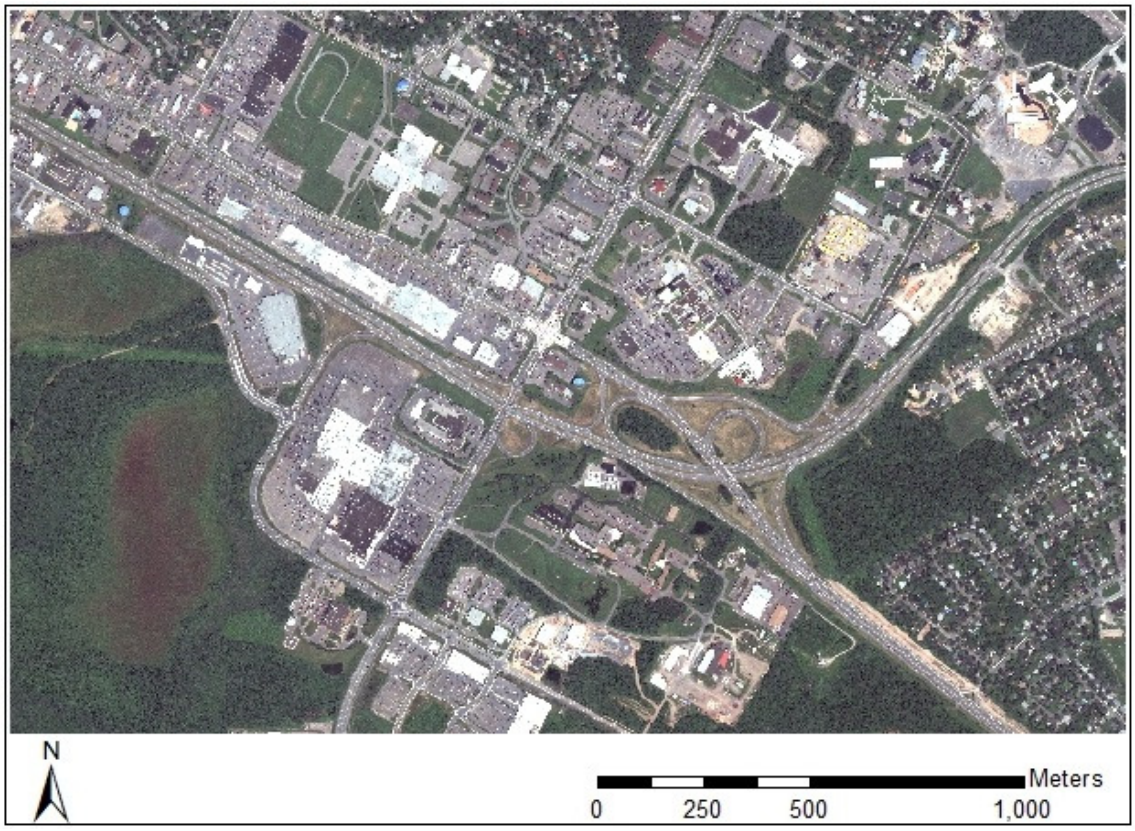


Figure 4.1 Pan-sharpened WorldView-2 image of the study area

4.3 Methods

Our proposed methodology consists of three major parts. First, the VHR image is segmented into different levels of segmentation resulting in several spectral and spatial features of objects. Second, discrete wavelet transform is applied to each set of spectral and spatial features at each object location to reduce the number of spectral and spatial features. Finally, the features extracted by the wavelet algorithm are classified using the

well-known maximum likelihood (ML) classification algorithm. The flowchart of the proposed method is shown in Figure 4.2. The remainder of this section describes each step involved in the proposed method.

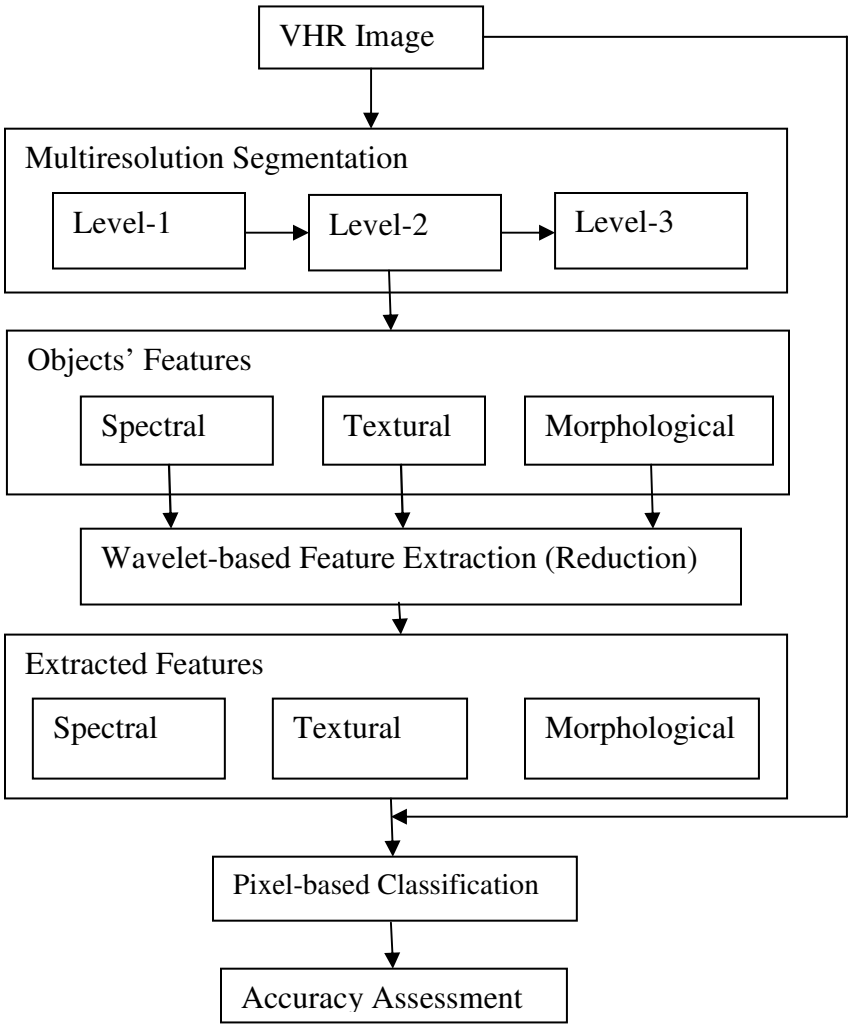


Figure 4.2 Flowchart of the proposed method

4.3.1 Image Segmentation

Image segmentation is the first step in object-based image analysis. One of the widely-used segmentation techniques for the analysis of urban areas is multiresolution

segmentation. This technique partitions the image into objects at different levels of segmentation. At higher levels, the size of objects increases and different classes of land cover can be better represented according to their size and spectral homogeneity at different levels. In this study, the multiresolution segmentation available in eCognition software was used. To avoid the trial and error process of selecting three parameters of multiresolution segmentation, namely scale, shape, and compactness [eCognition, 2011], the Fuzzy-based Segmentation Parameter optimizer (FbSP optimizer) [Tong *et al.,* 2012] was employed to get the optimal parameters for different levels of segmentation. FbSP optimizer is a supervised method which needs to be trained by selecting a few representative objects for each class at each level of segmentation. Then, taking into consideration the spatial and spectral features of training objects, FbSP gives the optimal parameters for each level [Zhang *et al.,* 2010; Tong *et al.,* 2012].

To use the FbSP optimizer, the first level of segmentation (L1) is carried out by manually selecting the three parameters. Often, the eCognition default values are used for shape and compactness, and the scale parameter is set in such a way that the resulting objects are smaller than the real world objects [Zhang *et al.,* 2010; Tong *et al.,* 2012]. Considering the spatial resolution of each image, the scale parameter was set to 20, 30, and 40 for WV-2, QB, and IK respectively. For the second level of segmentation (L2), the FbSP optimizer calculates the optimal parameters using objects in L1 and target (training) objects. This process is repeated until the desired number of segmentation levels is achieved. In this work, segmentation was stopped at the third level (L3) where the largest classes (e.g., large buildings and parking lots) are best outlined by objects.

Obviously in this level under-segmentation occurs for smaller classes (e.g., houses and single trees). Table 4.1 shows the parameters for three levels of segmentation of images. As mentioned, parameters in L1 were selected manually, but for L2 and L3 parameters were calculated by FbSP optimizer.

Table 4.1 Segmentation parameters for the three VHR images used in our study

Image	Level	Scale	Shape	Compactness	No. of objects
IK	L1	40	0.10	0.50	69,102
	L2	82	0.57	0.80	10,810
	L3	142	0.60	0.80	2,908
QB	L1	30	0.10	0.50	64,481
	L2	78	0.64	0.81	7,634
	L3	131	0.50	0.81	2,793
WV-2	L1	20	0.10	0.50	249,402
	L2	64	0.60	0.80	19,870
	L3	121	0.60	0.80	6,338

4.3.2 Spectral and Spatial Features

Having segmented the image, a large number of spectral and spatial features can be calculated for each object in eCognition. Three main feature types, spectral, textural, and morphological features, were investigated in this study. Table 4.2 shows the features calculated for each image. The description of all of these features can be found in eCognition [2011]. Due to the importance of spectral features for classification [Salehi *et al.,* 2012a], almost all spectral features available in eCognition are used in this study. These are 32 features for the WV-2 image and 19 features for QB and IK images for each level of segmentation.

Textural features calculated from the grey level statistics, especially those of grey level co-occurrence matrix (GLCM) have been used for a long time to improve classification

accuracy [Carleer *et al.,* 2006]. Among the fourteen GLCM texture measures, originally proposed by Haralick [1979], Entropy, Angular Second Moment, Homogeneity, and Contrast are the most frequently used texture features for urban land cover classification of VHR imagery in the literature [Salehi *et al.,* 2012b]. In this study, these four features are calculated for each band resulting in 32 features for WV-2 and 16 features for QB and IK in each level of segmentation. Morphological features (also known as Geometry in eCognition) are important for distinguishing spectrally similar classes such as buildings, parking lots, and roads. In eCognition these features are calculated based on the shape, extent, polygon, and skeleton of objects. For each image, 16 different morphological features were calculated for each level of segmentation (Table 4.2). Figure 4.3 shows examples of spectral, textural, and morphological features calculated for the objects in L3 of the WV-2 image.

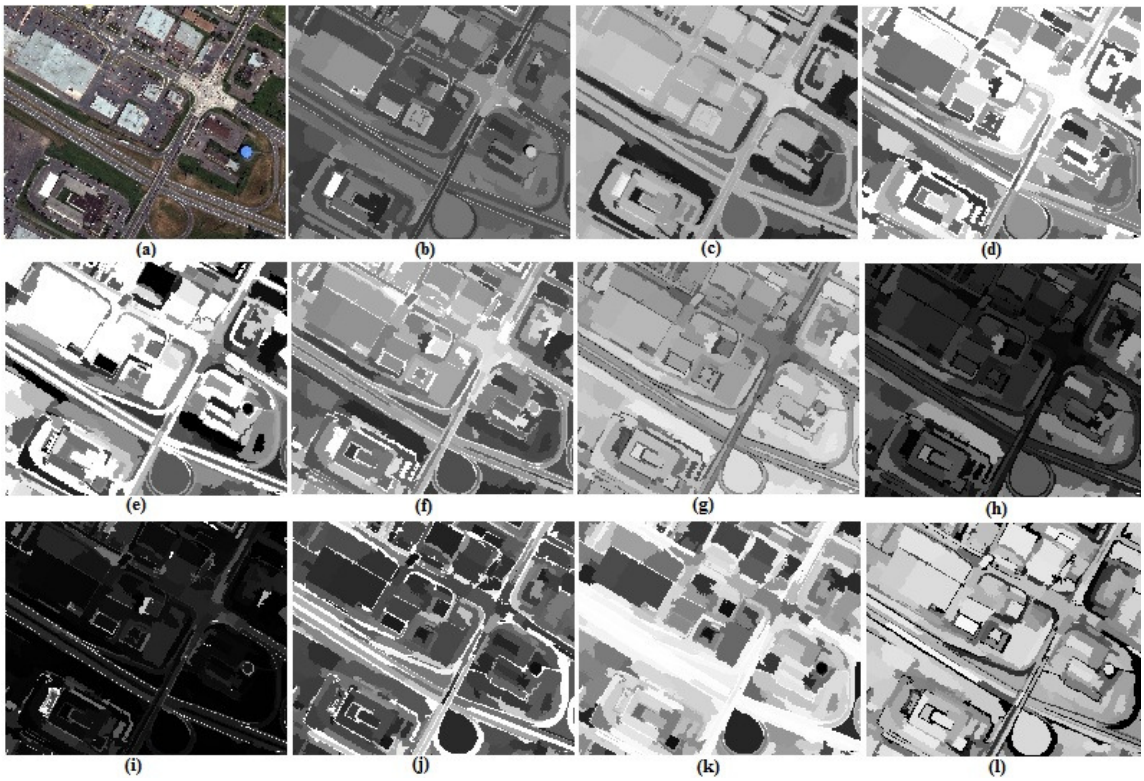


Figure 4.3 (a): pan-sharpened image, (b): Brightness, (c): Ratio R, (d): Standard deviation of RE, (e): Skewness of N2, (f): Entropy, (g): Homogeneity, (h): ASM, (i): Contrast, (j): Length/Width of skeleton, (k): Asymmetry, (l): Rectangular Fit.

Table 4.2 Three types of features calculated for each level of segmentation. The numbers in parentheses show the number of features for wv-2, QB, and IK images in turn (QB and IK have the same numbers).

Spectral **32 Features for WV-2** **19 Features for QB and IK**	Brightness(1)(1), Maximum difference(1)(1), Mean of all bands(8)(4) , Ratio of all bands(8)(4), Normalized difference vegetation index(NDVI)for N1and N2(2)(1), Standard Deviation of all bands (8)(4), Skewness of C, Y, RE, and N2 for WV-2 and of B, G, R, N for QB, and IK (4)(4)
Textural-GLCM **32 Features for WV-2** **16 Features for QB and IK**	Entropy(8)(4), Homogeneity(8)(4), Angular 2nd Moment (8)(4), Contrast(8) (4) of all bands
Morphological **16 Features for WV-2, QB, and IK**	Area, Length, Width, Asymmetry, Border Index, Density, Elliptic Fit, Radius of Largest Enclosed Ellipse, Radius of Smallest Enclosing Ellipse, Rectangular Fit, Roundness, Shape index, Average Length of Edges(polygon), Perimeter(polygon), Compactness(polygon), Length/Width(skeleton)

4.3.3 Wavelet-based Feature Extraction

Combining the total number of features for each type (spectral, textural, and morphological) creates high dimensional datasets. The dimensionality further increases when these features are combined with the original pan-sharpened bands of VHR imagery. For example, the combination of 32 spectral features of WV-2 (Table 4.2) with the 8 pan-sharpened bands of WV-2 results in a 40-dimension dataset. Because of the so-called "curse of dimensionality" [Bellman, 1961; Bishop, 1996], classification of such a high dimensional dataset will lead to inaccurate results. Curse of dimensionality arises when the data dimension (e.g. the number of objects' features forming a high dimensional dataset) increases resulting in a sparse dataset. Consequently, in order to have a reliable and sound statistically analysis (e.g. classification), the amount of data needed to estimate the statistics (e.g. training data) often increases exponentially. Therefore, feature reduction (or extraction) is a necessary step before the classification of such high dimensional datasets. The feature extraction employed in this study is based on discrete wavelet transforms.

Wavelet transforms are being used in the remote sensing community for applications such as registration, fusion, and compression. It has also been used for dimensionality reduction of hyperspectral data [e.g., Bruce *et al.,* 2002; Kaewpijit *et al.*, 2002]. In this study, wavelet transforms are employed to reduce the number of object features resulting from segmenting the image. Wavelet transform of a signal f(x), or an image f(x,y), is the inner product of the dilated and translated version of a basic function, also known as the "mother wavelet", with f(x):

$$\psi_{a,b}(x) = |a|^{-1/2} \psi\{(x-b)/a\} \quad (1)$$

$$\tilde{f}(a,b) = < f(x), \psi_{a,b}(x) > = \int_{-\infty}^{+\infty} f(x).\psi_{a,b}(x) \quad (2)$$

Where $\psi(x)$ and $\psi_{a,b}(x)$ are the mother wavelet and its dilated and translated version, also known as the "daughter wavelet", respectively. The < > sign denotes inner product and $\tilde{f}(a,b)$ is the wavelet coefficient. In discrete wavelet transform, the dilation (a) and translation (b) values are a discrete lattice of points (Embrechts *et al.,* 1995). A standard choice of a and b is by dyadic sampling such that $a = 2^{-j}$ and $b = k(0 \le k < 2^j)$, where j and k are integers.

The mother wavelet is regarded as a high pass filter which covers the high frequencies in the signal. In this paper, the multiresolution algorithm [Mallat, 1989] is utilized to compute wavelet transforms. For multiresolution wavelets, a low pass filter is needed to cover the low frequencies. This low pass filter, denoted by $\varphi(x)$, is called the "father wavelet" and its dilated and translated version $\varphi_{a,b}(x)$ is known as the "son wavelet" [Amolins *et al.,* 2007]. In a 1-D multiresolution wavelet transform, the daughter and son wavelets are applied to the signal, decomposing it into the high frequency (details) and low frequency (approximation) parts. This is done recursively by reapplying the same procedure to the low frequency part (Figure 4.4). Because of the dyadic sampling, the size of signal becomes half at each level of decomposition resulting in a smoother signal at higher levels.

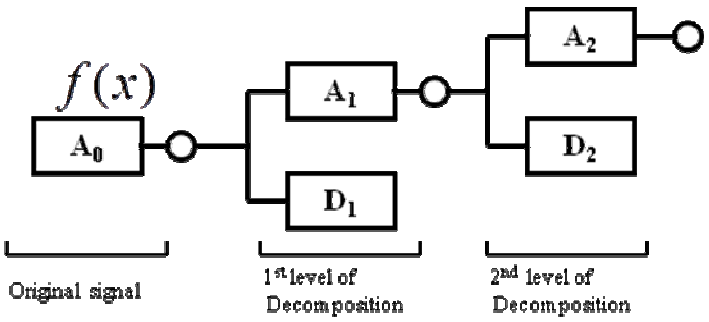


Figure 4.4 1-D multiresolution wavelet transform for two levels of decomposition. F(x) is the original signal and Ai and Di denote the approximation and detail parts of the signal at ith level of decomposition.

In this paper, a 1-D wavelet transform is applied to each object's signature for each set of spectral, textural, and morphological features. This procedure is repeated in the second level of decomposition and the approximation part is regarded as the new signature of the object. This process is done for all objects and the results are combined forming the extracted features of the image. Figure 4.5 provides a general illustration of the method. In order to keep the local characteristics of each set of features, only two decomposition levels of db2, one of the simplest and most localized members of the Daubechies family [Daubechies, 1992], were used.

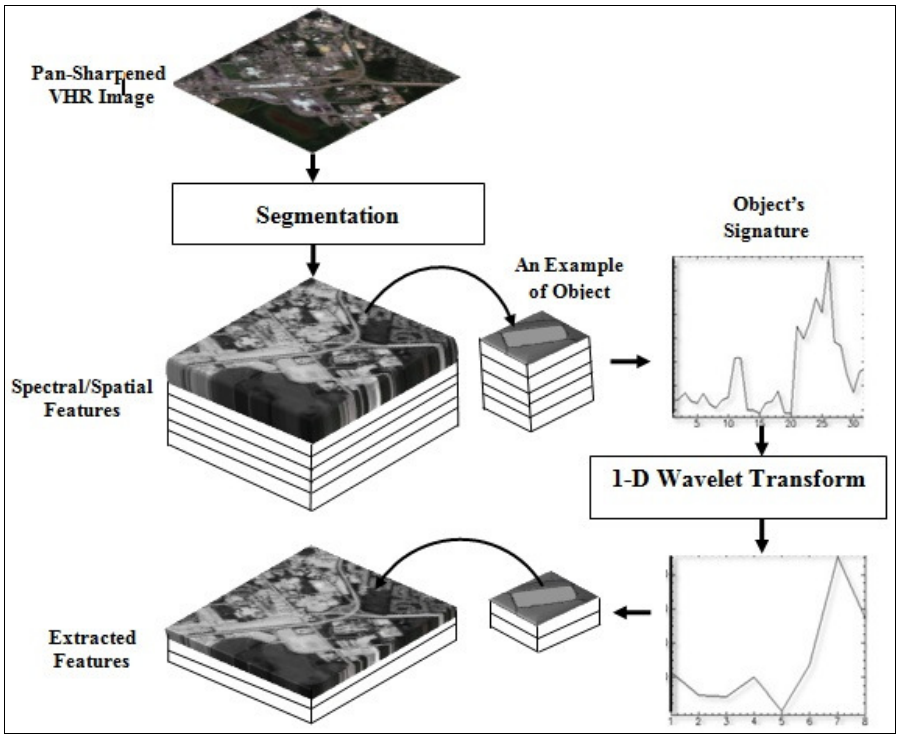
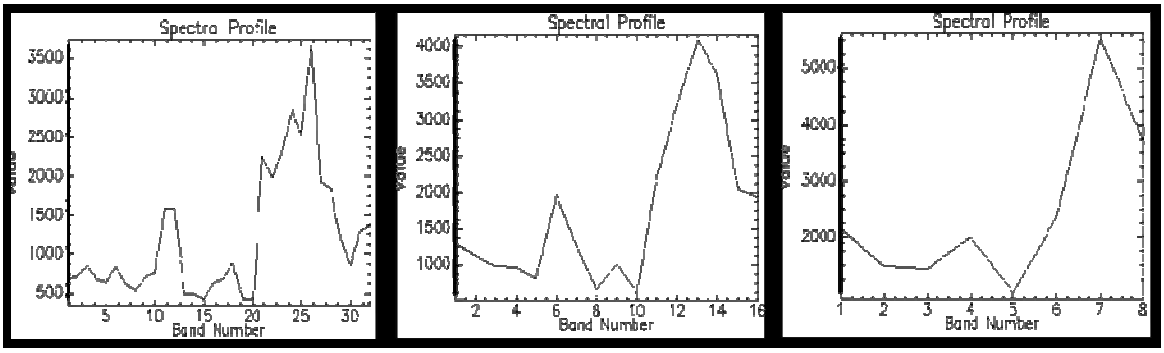


Figure 4.5 Workflow of the proposed wavelet-based feature extraction method

Figure 4.6 shows an example of the original spectral signature (32 features) of a building object in WV-2 image and its corresponding approximation for two levels of decompositions. In level 2, the number of features is reduced to eight from its original 32 features. Nevertheless, as can be seen, the overall structure (main peaks and valleys) is preserved. It should be noted that the 1-D wavelet transform is applied to the spectral/ spatial signature of objects rather than pixels. By doing that, the computation cost is hugely reduced since the total number of objects is considerably less than that of pixels in the image.



(a) (b) (c)

Figure 4.6 Example of the original spectral signature of a building object (a) and its corresponding signature after the first (b) and second level (c) of wavelet decomposition. The number of features in the second level is four times less than that of the original signature. Nonetheless, the overall structure of the signature is still preserved.

4.3.4 Classification

After the application of wavelet-based dimensionality reduction to each set of features, the extracted features together with the original bands of the image are employed in the classification process. To assess the effectiveness of each individual set of spectral, textural, and morphological features, the classification was performed on each set separately. That is, for each image, the extracted features of each type (spectral, textural, and morphological) are stacked to the original pan-sharpened bands creating three different datasets for each level of segmentation (L2 and L3) followed by classification. Maximum Likelihood (ML) classification, is the most commonly classifier with remote sensing data [Riachrds and Jia, 2006], and are served as the reference classifier in most related literature [Chini *et al.*, 2008].

4.4 Results

4.4.1 Feature Extraction Results

Table 4.3 reports the number of spectral, textural, and morphological features extracted by wavelet transforms for the three VHR images. For each of WV-2, QB, and IK, two levels of decomposition of the 1-D wavelet transform were applied to the spectral and spatial (i.e., textural and morphological) signature of each individual object. To demonstrate the effect of objects' size (segmentation levels) on classification accuracy, feature reduction was carried out on objects in L2 and L3 independently. The result for each level was then used in the subsequent classification. For the N number of original features, the number of extracted features at the L^{th} level of wavelet decomposition is $N/2^L$ because of the dyadic sampling of the algorithm. For example, for 32 spectral features of WV-2 (Table 4.2), wavelet transform results in 8 features in the second level of decomposition. The extracted features are linear and weighted combinations of the original features. Multiresolution wavelet works perfectly when the number of input features is an integer power of two or at least even. Since the number of spectral features for IK and QB is 19, I simply repeated the 19^{th} feature to increase the number of features to 20. For these 20 features wavelet transforms extract five features at the second decomposition level.

Table 4.3 Numbers of features resulting from wavelet feature extraction at the second level of decomposition

Feature name	WV-2	QB	IK
Spectral	8	5	5
Textural-GLCM	8	4	4
Morphological	4	4	4

4.4.2 Classification Results

The results of ML classification in the form of the conventional error matrix [Congalton and Green, 2009] and its elements including the overall accuracy (OA), kappa coefficient (KC), producer's accuracy (PA), and user's accuracy (UA) are presented and discussed in this section. In some cases the average PAs of buildings, roads, and parking lots are presented as the PA of impervious classes (Imp.). Similarly, the average PAs of grass and trees demonstrates the PA of vegetation (Veg).The results are summarized in four groups. First, the classification accuracies of all datasets are presented and discussed. Secondly, the effectiveness of wavelet feature extraction is demonstrated and compared with the conventional principal component analysis (PCA). Then, lastly, the effects of types of features and levels of segmentation on classification are investigated.

All the three VHR images with their corresponding spectral and spatial features are employed in each test. ML classification is a supervised method and needs to be trained by sample data. To evaluate the performance of the proposed framework on the limited number of training samples, only about 5% of the study area of in each image is used for training the classifier. On the other hand, for more reliable validation of results, a large part of the study area (about 60%) of each image is used to test the classification results. Table 4.4 shows the details of training and test areas for each image. Although the number of training and testing pixels is different among the three images, the training and testing areas are close for them. This is because of the different in pixel size between the three images.

Table 4.4 Number of training and test pixels used in classification and Accuracy assessment, respectively

Class Name	WV-2		QB		IK	
	Training data: No. of pixels	**Testing data: No. of pixels**	**Training data: No. of pixels**	**Testing data: No. of pixels**	**Training data No. of pixels**	**Testing data No. Of pixels**
Road	72748	1302755	42583	553883	19993	284495
Building	141802	1604169	88428	638052	43823	336127
Parking lot	154384	2020212	94333	657978	45190	372773
Grass	287480	3114281	147548	1254572	63829	489789
Tree	183560	3812203	144584	2113810	61998	976584
Total	839974	11853620	517476	5218295	234833	2459768
% of the image	**4.7%**	**66.5%**	**5.7%**	**57.4%**	**5.3%**	**55.2%**

4.4.2.1 Classification Accuracies of All Datasets

As mentioned earlier, the extracted spectral, textural, and morphological features by wavelet transforms in both levels of segmentation were stacked to the original pan-sharpened bands of each image and ML classification is applied to each dataset. The classification accuracies for each dataset corresponding to WV-2, QB, and IK are listed in Table 4.5, 4.6, and 4.7, respectively. Each table contains the classification accuracies provided by the original pan-sharpened bands (PS), and the stacked set of PS bands and spectral, textural , and morphological features of objects in L2, L3, and the combination of L2 and L3 (L2+L3).

Table 4.5 ML classification accuracies of the WV-2 image for different sets of objects' features and for two levels of segmentation extracted by wavelet. The number in parentheses represents the number of features which are stacked to the 8 PS bands for classification.

Class Name	8PS	8PS+Spectral			8PS+ Textural			8PS+ Morphological		
		L2 (8)	L3 (8)	L2+L3 (16)	L2 (8)	L3 (8)	L2+L3 (16)	L2 (4)	L3 (4)	L2+L3 (8)
	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA
Road	38.8 46.2	54.7 60.9	58.8 67.1	63.5 62.1	54.6 45.7	61.7 65.2	61.4 53.2	83.9 70.2	69.9 83.8	83.3 68.8
Building	52.7 63.4	79.4 66.8	85.4 74.2	79.2 76.3	42.2 84.9	52.2 71.1	44.4 79.6	68.1 63.4	79.2 54.1	57.1 68.1
Parking lot	57.8 41.2	64.0 70.9	71.0 75.7	70.7 78.3	68.6 25.0	66.5 51.3	66.1 37.1	68.2 80.4	69.1 78.2	80.3 74.2
Grass	85.5 71.2	94.1 83.2	94.7 78.5	95.3 79.4	93.8 85.2	94.2 81.9	94.0 84.1	69.1 73.1	86.3 66.7	66.7 71.8
Tree	80.2 91.3	89.4 96.2	86.3 97.0	87.4 97.1	91.0 95.5	89.3 95.4	91.4 94.8	93.3 93.0	77.1 92.1	92.8 94.5
OA	68.7	80.7	82.2	82.7	74.0	77.8	75.7	78.2	77.0	77.9
KC	59.6	74.8	76.7	77.4	66.3	71.1	68.4	71.1	69.9	71.1

Table 4.6 ML classification accuracies of the QB image for different sets of objects' features and for two levels of segmentation extracted by wavelet. The number in parentheses represents the number of features which are stacked to the 4 PS bands for classification.

Class Name	4PS	4PS+Spectral			4PS+ Textural			4PS+ Morphological		
		L2 (5)	L3 (5)	L2+L3 (10)	L2 (4)	L3 (4)	L2+L3 (8)	L2 (4)	L3 (4)	L2+L3 (8)
	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA
Road	43.0 78.6	53.5 74.3	77.5 77.5	56.8 69.8	50.9 79.0	50.0 83.1	48.8 81.7	80.0 84.8	75.1 88.0	78.7 87.4
Building	72.7 63.8	63.8 80.9	84.2 84.2	67.4 85.4	77.1 59.1	78.2 60.7	73.6 60.2	62.1 72.8	60.5 63.4	59.6 66.9
Parking lot	53.1 26.9	65.7 36.7	50.5 50.5	67.7 51.9	53.7 48.9	51.9 46.0	53.2 46.4	60.4 53.4	59.7 60.0	59.1 59.7
Grass	85.7 86.2	94.1 90.6	92.3 92.3	93.6 89.8	96.1 89.9	93.0 89.4	94.1 87.9	86.6 86.7	83.7 88.0	83.2 87.4
Tree	94.0 91.2	97.6 95.3	93.8 93.8	98.5 94.2	97.0 96.1	98.2 93.6	97.8 81.7	94.6 91.5	95.1 88.1	95.0 87.4
OA	77.2	82.8	85.1	84.2	82.3	81.4	80.9	82.5	80.8	80.7
KC	69.9	76.8	79.9	78.7	76.1	75.0	74.3	76.4	74	74.1

Table 4.7 ML classification accuracies of the IK image for different sets of objects' features and for two levels of segmentation extracted by wavelet. The number in parentheses represents the number of features which are stacked to the 4 PS bands for classification.

Class Name	4PS	4PS+Spectral			4PS+ Textural			4PS+ Morphological		
		L2 (5)	L3 (5)	L2+L3 (10)	L2 (4)	L3 (4)	L2+L3 (8)	L2 (4)	L3 (4)	L2+L3 (8)
	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA	UA PA
Road	42.0 70.0	51.9 66.9	59.6 57.1	64.1 55.4	60.2 65.8	55.2 72.2	68.9 62.7	55.8 80.5	80.0 79.9	77.9 82.6
Building	69.4 65.3	76.0 70.6	64.9 72.9	63.9 76.0	75.0 70.0	66.9 72.4	63.1 68.2	39.9 56.0	65.7 64.1	74.0 60.5
Parking lot	56.1 26.3	66.3 53.0	74.0 66.2	71.9 64.3	62.2 61.9	58.4 48.3	61.5 69.7	25.8 60.8	62.7 61.2	63.5 68.4
Grass	73.5 95.9	83.1 97.0	83.5 97.7	83.9 97.6	82.9 95.8	81.6 93.1	78.9 92.7	57.1 96.3	76.2 96.0	78.0 95.6
Tree	96.6 84.6	97.9 91.3	98.3 90.9	98.2 91.6	98.2 90.3	99.0 86.9	99.0 85.1	72.8 87.3	97.4 86.5	97.3 87.8
OA	73.7	81.0	82.2	82.3	81.5	78.6	79.4	80.4	80.7	82.1
KC	65.4	74.8	76.3	76.5	75.5	71.8	72.9	74.0	74.5	76.2

As expected, incorporating additional spectral and spatial features significantly increases the overall classification accuracy and the accuracy of individual classes for all three images. For example, for WV-2, adding 8 spectral features of objects in L3, extracted by the wavelet-based feature extraction method, increases the KC by about 17% compared with the case when only 8 PS bands are utilized in classification. Similarly, the additional 5 spectral features of objects in L3 increase the KC of QB and IK by 10% and 11%, respectively. Incorporating additional features of objects into classification makes a greater contribution to increasing the PA of three impervious land cover types (roads, buildings, and parking lots) than the PA of vegetation (grass and trees). Utilizing the extracted spectral features of objects in L3 increases the average PA of impervious classes by 20% for WV-2, and about 14% and 12% for QB and IK, respectively. For

vegetation, this increase is only about 6%, 4%, and 4% in WV-2, QB, and IK data. Figures 4.7, 4.8, and 4.9 show the classification results of the PS bands only and PS bands together with extracted spectral features of objects in L3 for WV-2, QB, and IK. As seen, in all three figures, the addition of object features into classification has significantly reduced misclassification of similar classes such as buildings, roads, and parking lots. The majority of impervious land cover classes in the image including small houses in the south-east part of the image have been correctly classified, especially for the WV-2 image. In addition, as seen in these figures, the incorporation of object features avoids the so-called salt and pepper effect, caused by the spectral heterogeneity of classes, which is clearly seen for when only PS bands are utilized.

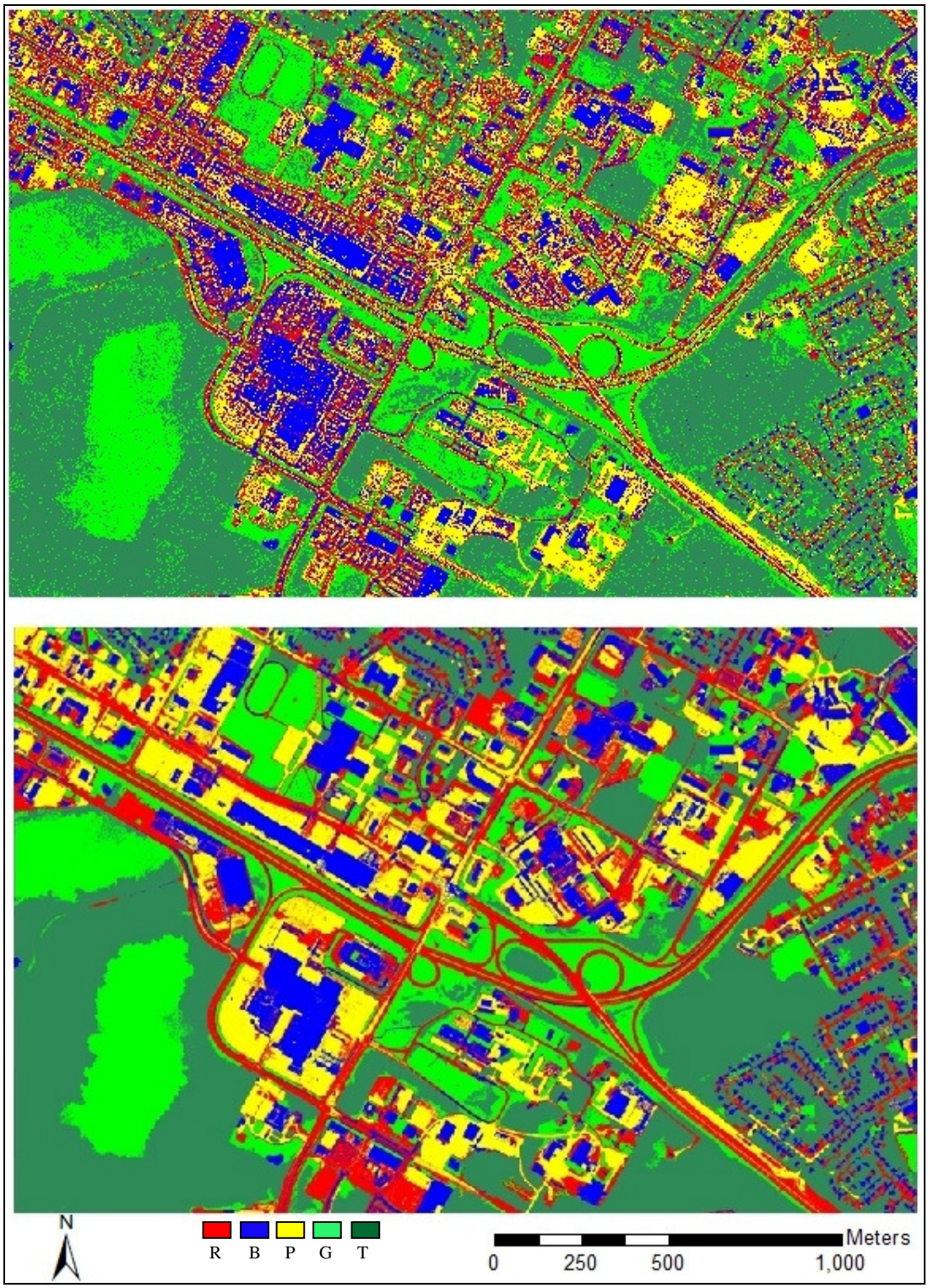


Figure 4.7 ML classification result of (a) 8 PS bands and (b) 8 PS+8 Spectral features for the WV-2 image. R: roads; B: buildings; P: parking lots; G: grass; T: trees.

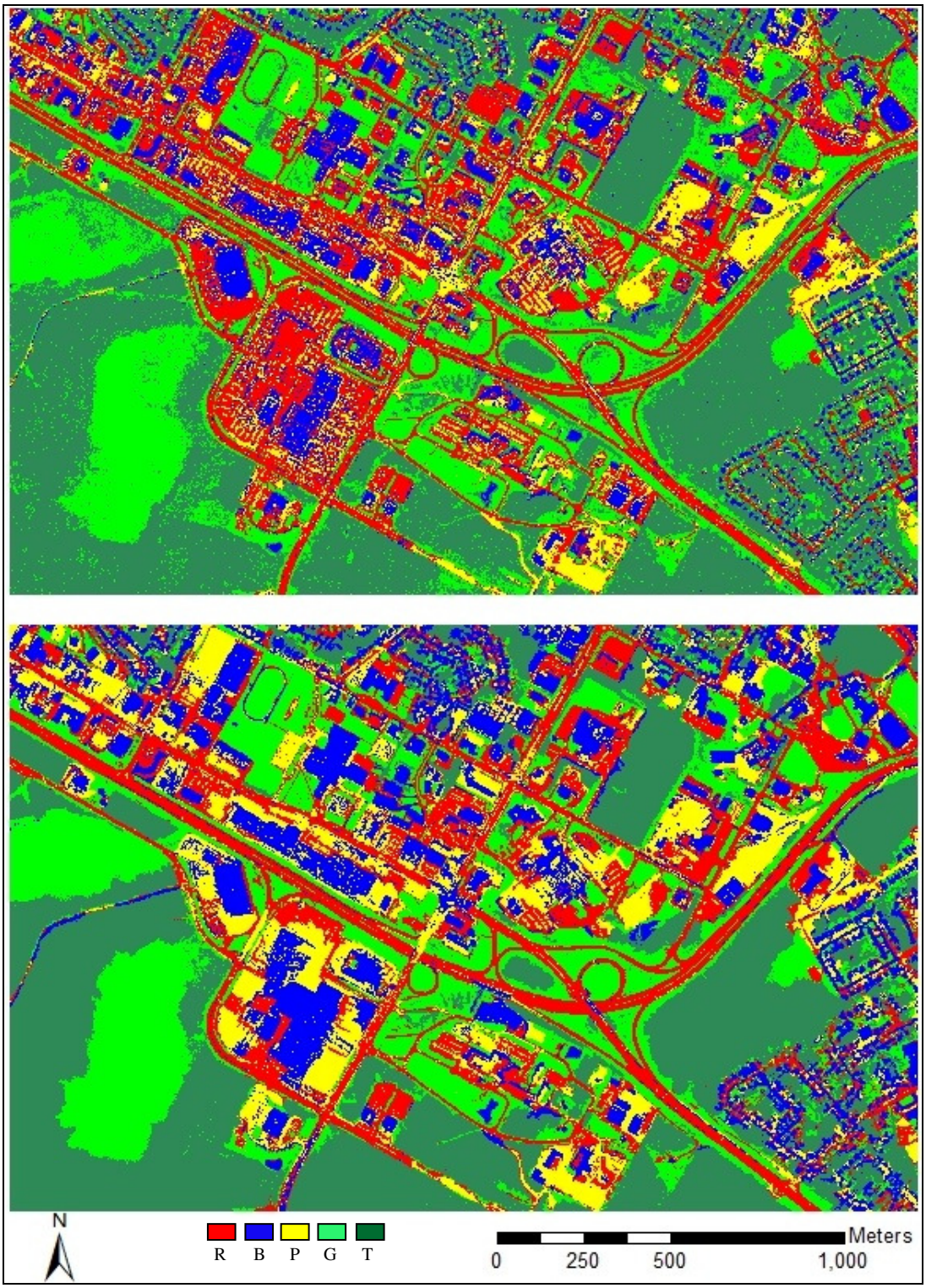


Figure 4.8 ML classification result of (a) 4 PS bands and (b) 4 PS+5 Spectral features for the QB image. R: roads; B: buildings; P: parking lots; G: grass; T: trees.

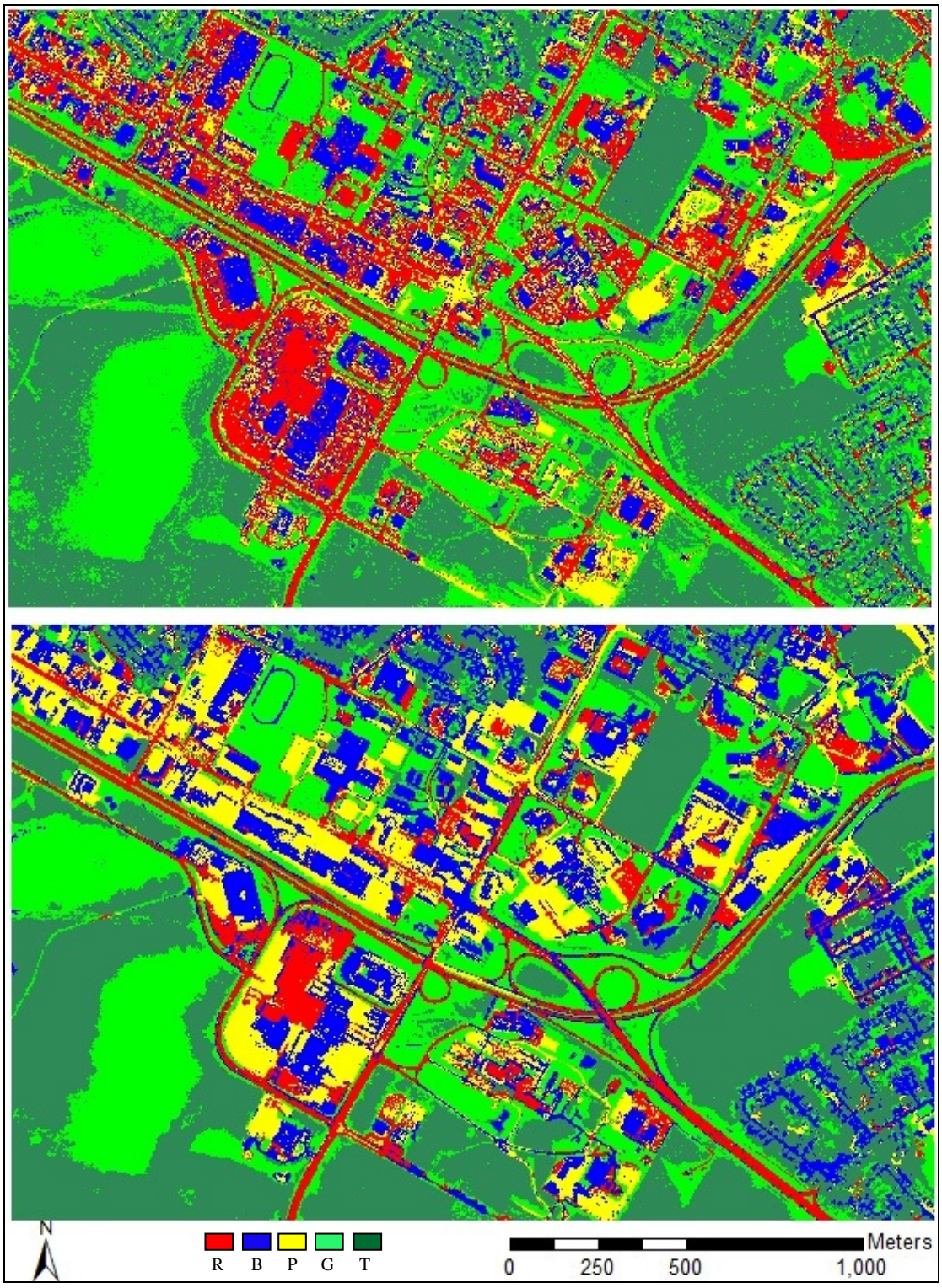


Figure 4.9 ML classification result of (a) 4 PS bands and (b) 4 PS+5 Spectral features for the IK image. R: roads; B: buildings; P: parking lots; G: grass; T: trees.

4.4.2.2 Type of Features

To examine the usefulness of each type of feature (spectral, textural, and morphological) in classification results provided by each type are plotted in graphs and shown in Figure 4.10. Despite the difference in spatial resolution of the three VHR images and the difference in number of spectral bands between WV-2 and the two other images, very similar trends are seen in these three graphs, especially for WV-2 and QB. This shows there is little dependency of our method on the input VHR imagery.

For all the three images, spectral features give the highest OA among all the three types of features. This confirms the importance of incorporating the spectral information, rather than spatial information, of neighboring pixels (in the form of objects) into the classification of VHR imagery. Spectral information particularly has a more significant impact on in increasing the classification accuracy of vegetation which in turn increases the overall accuracy of the image.

Morphological features provide comparable (for WV-2 and QB) or better (for IK) average PA of impervious class (Imp) compared to textural and spectral features. This result is not surprising as buildings, roads, and parking lots have very similar texture and spectral information, but they are distinguishable by their morphological features such as size, shape, and skeleton. On the other hand, looking at Tables 4.5, 4.6, and 4.7, incorporating morphological features into classification lowers the PA of vegetation compared to using only PS bands. Grass and tree areas are normally very diverse in terms of their shape and size and thus morphological features are not suitable representatives of these classes.

Textural features make the least contribution to increasing the accuracy of impervious classes, while for vegetation texture plays an important role. This is because of not only the textural similarity between buildings, roads, and parking lots, but also the textural diversity within each impervious land cover. For example, there are parking lots with cars (commercial) and without cars (residential) creating different textural characteristics for the same class of parking lots.

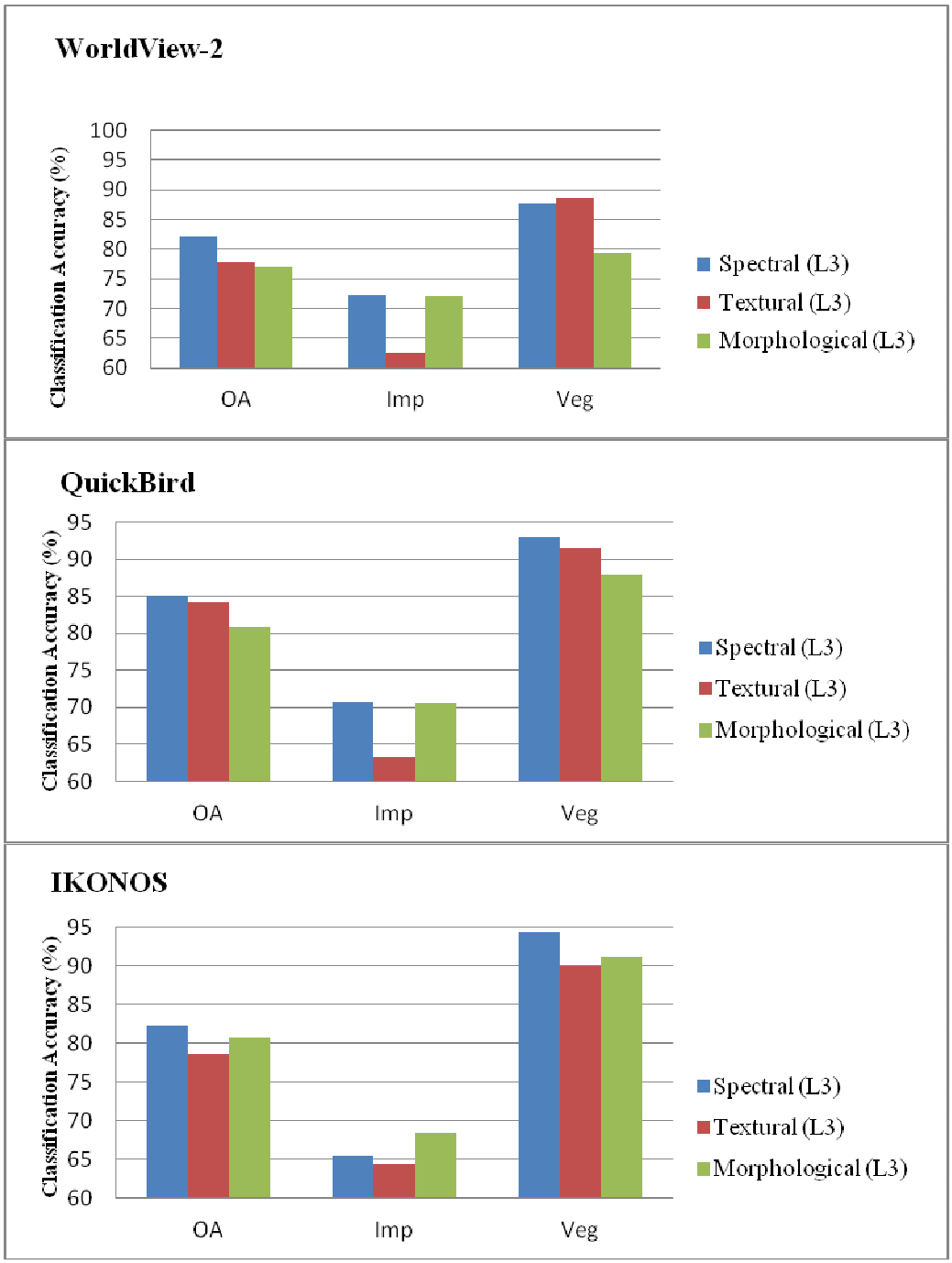


Figure 4.10 Comparison of classification accuracies achieved by each set of spectral, textural, and morphological features extracted by wavelet transforms.

4.4.2.3 Level of Segmentation

To demonstrate the role of the level of segmentation, classification accuracies of each individual level as well as the combination of both levels are considered. Figure 4.11 summarizes the classification accuracy of different types of features in different levels of segmentation for all the three images. Three sets of classification accuracies are plotted for each of OA, Imp, and Veg. One set is for the case when objects in the second level of segmentation are utilized (L2). The second set is for the objects in the third level of segmentation (L3), and the last set is the combination of objects in both levels (L2+L3). As can be seen, the combination of two levels does not yield significant improvement over the case when only one level of segmentation is utilized. This suggests the use of features from only one level of segmentation instead of both levels. Consequently, the computation costs decrease significantly if only one level, instead of two, is utilized.

Of L2 and L3, the latter results in a better classification accuracy for impervious land covers for almost all 9 cases shown in Figure 4.11. This result can be explained by the fact that the three impervious land cover types are better represented by larger objects (L3) than small ones (L2) and the FbSP segmentation method results in more meaningful objects for these three classes in L3 rather than in L2. However, for the class of vegetation, L2 slightly outperforms L3. Grass and tree areas typically do not possess specific shape and thus over-segmentation, which occurs in the lower level of segmentation (e.g., L2), does not have a significant effect on classification of them. Further, for vegetation, the classifier mainly relies on the spectral information of individual pixels which are better preserved in the lower level of segmentation. In other

words, the spectral information of objects in *lower* levels, rather than higher levels, is closer to the original spectral information of the individual pixels.

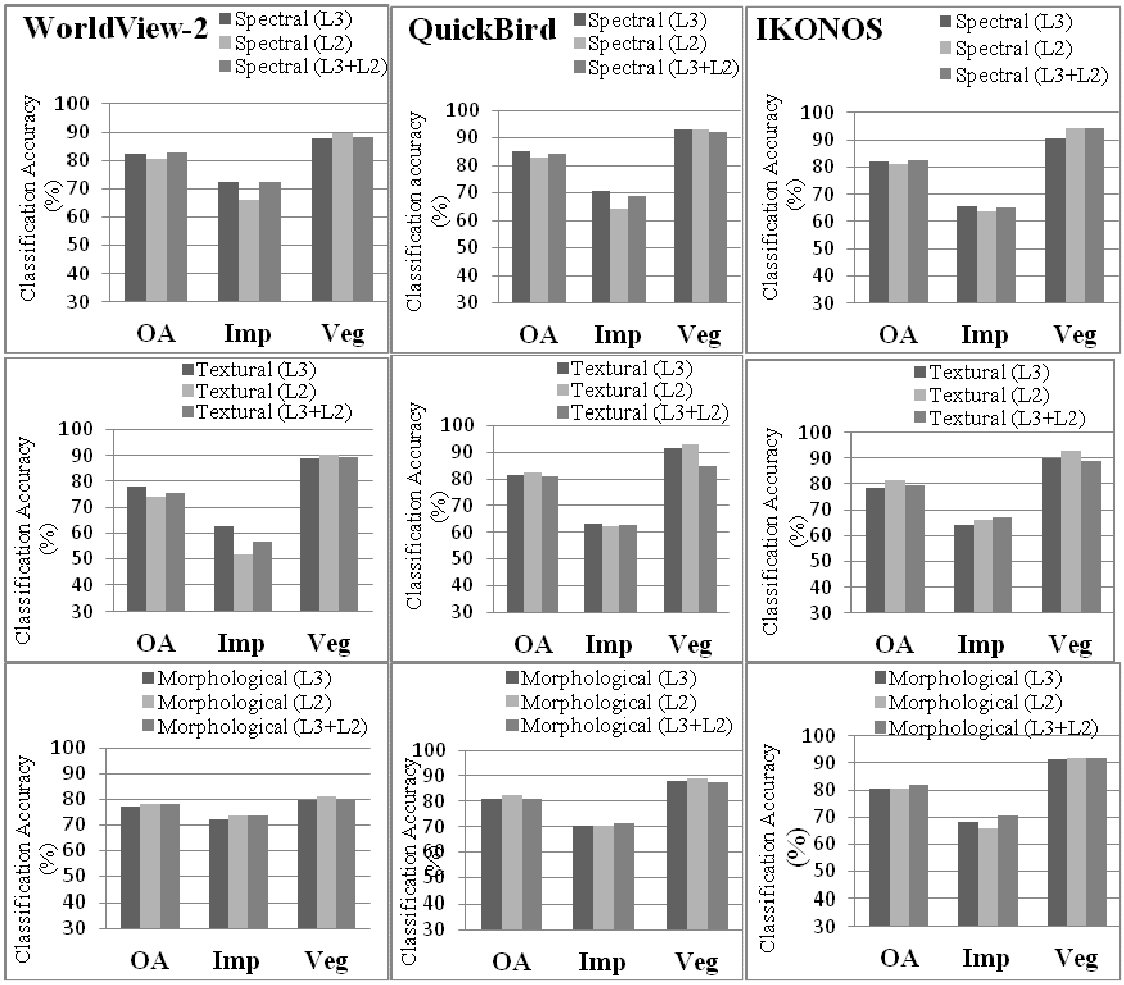


Figure 4.11 Comparison of the classification accuracies achieved by objects' feature in L2, L3, and the combination of L2 and L3.

4.4.2.4 Wavelet vs. PCA

A standard method for dimensionality reduction is PCA. To compare the performance of PCA and wavelet feature extraction, the classification results of PCA and wavelet

methods are compared. PCA is applied to the spectral features of objects in L3 for all three images. Then, for the WV-2 image, the first 8 PCs, corresponding to 8 wavelet-based extracted features, are stacked to the 8 PS bands creating 16 features. Similarly, for the QB and IK, the first 5 PCs are stacked to the PS bands creating 9 features for each of QB and IK images. ML classification is carried out on each dataset and the results are compared to the corresponding results achieved by wavelet. Figure 4.12 shows the accuracy provided by PCA and wavelet transforms. As can be seen from Figure 4.12, wavelet and PCA basically provide similar results in terms of classification accuracy. However, this study shows that wavelet transforms have a much higher computing efficiency when compared to PCA. Unlike the wavelet approach which is applied to each individual object, PCA takes into consideration all the individual pixels and performs pixel-by-pixel transformation of the image. Since the number of objects is significantly lower than the number of pixels in VHR imagery, especially in higher levels of segmentation (e.g., L3), the object-based wavelet feature extraction is much faster than the pixel-based PCA. By comparing the computational complexity of wavelet and PCA, it is found, in this study, that wavelet is dramatically less complex than PCA as the number of pixels is far greater than the number of objects forming the images, especially for VHR imagery. For N number of original features, the wavelet-based feature extraction method yields $O(N)$ computations per unit [Daubechies *et al.*, 1992] resulting in a total complexity of $O(MN)$ where M is the number of an image's objects. For PCA the total estimated time complexity is $O(LN^2 + N^3 + RLN)$ where L is the number of an image's

pixels and R is the number of formed component [Kaewpijit *et al.,* 2003; Chen and Qian, 2011]. For example, for 32 spectral features of WV-2, the computation complexity of wavelet transforms is on the order of $O(32 \times 6338) = O(202816)$, where 6338 is the number of objects in L3 (Table 4.1). For PCA, the computational complexity of 32 spectral features is on the approximate order of $O(3.6e + 10)$.

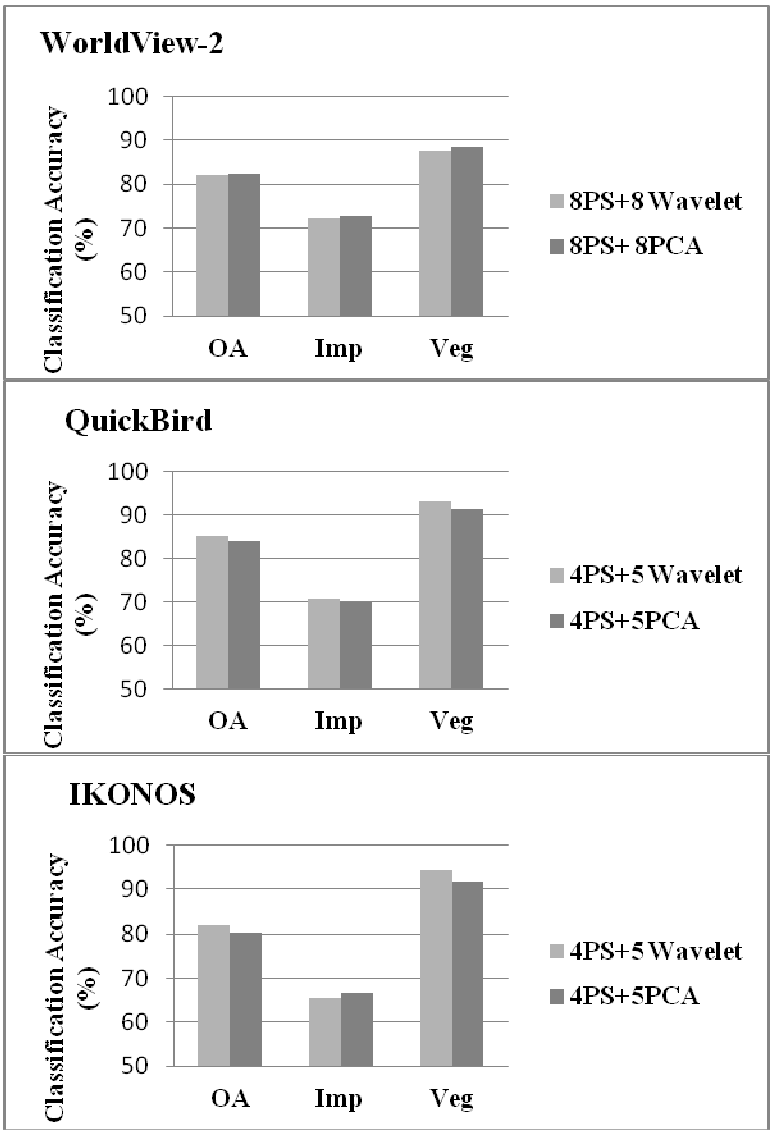


Figure 4.12 Comparison of classification results achieved by PCA- and Wavelet- based extraction of spectral features in L3. OA: overall accuracy; Imp: average producer's accuracies of impervious land covers; Veg: average producer's accuracies of grass and trees.

4.5 Conclusion

This study demonstrates a combined pixel- and object-based image analysis framework for urban land cover classification of VHR imagery. The method is tested on the WorldView-2, QuickBird, and IKONOS images of the same area and similar results are achieved for all of these three images. The proposed framework takes advantage of both approaches (pixel-based and object-based) by incorporating the spectral and spatial features of objects, resulting from segmentation, into automatic pixel-based classification. After automatic segmentation of the image using FbSP optimizer, tens of spectral, textural, and morphological features of objects are available. However, utilizing all these features creates huge computational costs and more importantly results decline in classification accuracy due to the "curse of dimensionality." To conquer the problem of dimensionality, a 1-D wavelet-based feature reduction/ extraction method was proposed and applied to each set of spectral, textural, and morphological features of objects. Finally, the conventional Maximum Likelihood classification was used to classify the combination of the original bands of the image and the wavelet extracted features.

Regardless of the type of features and level of segmentation, the classification results confirm that incorporation of additional spectral and spatial features of objects significantly improves the classification results, especially for the WV-2 image. Furthermore, the results demonstrate the usefulness of the wavelet-based feature

reduction method in terms of both the accuracy and computational complexity in contrast to the conventional pixel-based feature extraction algorithms such as PCA, the wavelet method is applied to each individual object (a group of pixels) rather than individual pixels, and thus the computation cost is dramatically reduced. Further, since wavelet transforms preserve the local characteristics of objects' signatures in lower dimension, they are more desirable for dimensionality reduction of larger numbers of features than are conventional feature extraction methods.

Of three types of objects' features, spectral features prove to be more effective for increasing the overall classification accuracy and the classification accuracy of individual classes of all the three images. In addition, for impervious land cover types such as buildings, roads, and parking lots, the morphological features show results comparable to those of spectral features. Impervious land covers normally possess specific shape and size and thus the morphological features (shape, size, skeleton, etc) of objects represent them well. On the other hand, impervious land covers are very diverse in terms of their textural properties. Consequently, textural features do not yield promising results for such classes as the quantitative results of this study indicate.

The multiresolution segmentation yields objects with different sizes in each level of segmentation which is well suited to urban areas containing classes with different sizes. Consequently, the same number of features can be extracted in each level of segmentation, and incorporated into classification. It has been found, in this study that utilizing the features of only one level of segmentation is sufficient. In other words, the combination of features from two levels does not increase the classification accuracy

compared to the use of only one level. Further, of L2 and L3, the latter shows better classification results. Utilizing objects in higher level of segmentation (L3) also requires much less computation time in the wavelet-based feature extraction method as the number of objects is significantly less in L3 compared to L2.

In this study only the spectral and spatial information of the image are extracted and incorporated in the classification process. Incorporation of ancillary data such as digital surface/elevation model data extracted from stereo VHR imagery would help the classifier to better distinguish between different impervious land cover types. This will be the direction of our future research.

Acknowledgements

The authors would like to thank the City of Fredericton for the QuickBird and IKONOS images used in this work. The WorldView-2 image was provided by DigitalGlobe® Inc. to the authors through "ERDAS IMAGINE- DigitalGlobe Geospatial Challenge" contest. The authors are grateful to DigitalGlobe for making such a high quality dataset available. This research is supported by the NSERC (Natural Science and Engineering Research Council of Canada) Strategic Project Grants and Discovery Grants Programs.

REFERENCES

Amolins, K., Y. Zhang, P. Dare, 2007.Wavelet based image fusion techniques- an introduction, review and comparison. *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (4): 249-263.

Bellman, R., 1961. *Adaptive Control Processes: A Guided Tour*, Princeton University Press, Princeton, New Jersey, 255 p.

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

Bishop C. 1996. *Neural Networks for Pattern Recognition*, Oxford University Press, New York, 1996, 482 p.

Bruce, L.M., C.H. Koger, and J. Li, 2002. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction, *IEEE Transactions on Geoscience and Remote Sensing*, 40(10): 2331-2338.

Chan, J. C., R. Bellens, F. Canters, and S. Gautama, 2009. An Assessment of geometric activity features for per-pixel classification of urban man-made objects using very high resolution satellite imagery. *Photogrammetric Engineering and Remote Sensing*, 75(4):397-411.

Chen, G., and S.E, Qian, 2011. Denoising of hyprespectral imagery using principal component analysis and wavelet shrinkage, *IEEE Transactions on Geoscience and Remote Sensing*, 49(3): 973-980.

Chini, M., F. Pacifici, W. J. Emery, N. Pierdicca, and F.D. Frate, 2008. Comparing statistical and neural network methods applied to very high resolution satellite images showing changes in man-made structures at Rocky Flats, *IEEE Transactions on Geoscience and Remote Sensing*, 46(6):1812-1821.

Congalton, R.G., and K. Green, 2009. *Assessing the accuracy of remotely sensed data: principles and practices,* Second edition. Taylor & Francis Group, Boca Raton, Florida, 183 p.

Daubechies I, 1992. *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 357 p.

DigitalGlobe 2009. *DigitalGlobe core imagery products guide* [online]. Available from: http://www.digitalglobe.com/index.php/6/DigitalGlobe+Products (last date accessed: 07 June 2012).

Duda, R.O., P.E. Hart, and D.G. Stork, 2001. *Pattern Classification*, Second edition, Wiley, New York, 680 p.

eCognition, 2011 . *eCognition Developer (8.7) Reference Book*. Trimble Germany GmbH, Trappentreustr. 1, D-80339 München, Germany.

Fukunaga, K., 1990. *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, 591 p.

GeoEye, 2009. *GeoEye product guide v1.0.1* [online]. Available from: http://www.geoeye.com/CorpSite/assets/docs/brochures/GeoEye_Product_Guide.pdf (last date accessed: 07 June 2011).

Haralick, R. M., 1979, Statistical and Structural Approaches to Texture. *In Proceedings of IEEE*, 67(5):786-804.

Huang, X., L. Zhang, and W. Gong, 2011. Information fusion of aerial images and lidar data in urban areas: vector-stacking, re-classification and post-processing approaches, *International Journal of Remote Sensing,* 32(1):69-84.

Kaewpijit, K. , J. Le Moigne, and T. El-Ghazawi, 2003. Automatic reduction of hyperspectral imagery using wavelet spectral analysis, *IEEE Transaction on Geoscience and Remote Sensing*, 41(4): 863–871.

Landgrebe, D. A., 2003. *Signal Theory Methods in Multispectral Remote Sensing*, Wiley, Hoboken, New Jersey, 528 p.

Lee, C. and D.A. Landgrebe, 1993. Feature extraction based on decision boundaries, *IEEE Transaction of Pattern Analysis and Machine Intelligence*, 15(3): 388-400.

Lu, D., S. Hetrick, and E. Moran, 2010. Land cover classification in a complex urban-rural landscape with Quickbird imagery. *Photogrammetric Engineering and Remote Sensing* 76(10):1159-1168.

Mallat, S. G., 1989. A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Pattern Analysis Machine Intelligence*, 11(7): 674–693.

Richards, J. A., and X. Jia (2006). *Remote Sensing Digital Image Analysis, An Introduction*, Fourth edition, Springer, Berlin, 464 p.

Salehi, B., Y. Zhang, and M. Zhong, 2012a. The effect of four new multispectral bands of WorldView2 on improving urban land cover classification, *Proceedings of ASPRS 2012 Annual Conferenc*e, 19-23 March, Sacramento, California(American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland), unpaginated CD-ROM.

Salehi, B., Y. Zhang, M. Zhong, and V. Dey , 2012b. A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery. *International Journal of GeoInformatics,*8(3):35-51.

Tong, H., T. Maxwell, Y. Zhang, and D. Vivek, 2012. A supervised and fuzzy-based approach to determine optimal multiresolution image segmentation parameters. Accepted for publication, *Photogrammetric Engineering and Remote Sensing.*

Zhang, Y., T. Maxwell, H. Tong, and V. Dey, 2010. Developement of supervised software tool for automated determination of optimal segmentation parameters for eCognition. *ISPRS TC VII symposium-100 Years ISPRS*, July 5-7, Vienna, Austria.

Chapter 5 : AUTOMATIC MOVING VEHICLES INFORMATION EXTRACTION FROM SINGLE-PASS WORLDVIEW-2 IMAGERY[1]

Abstract

Because of the sub-meter spatial resolution of very high resolution (VHR) optical satellite imagery, vehicles can be identified in this type of imagery. Further, because there is a time lag in image collection between the panchromatic (Pan) and multispectral (MS) sensors onboard VHR satellites, a moving vehicle is observed by the satellite at slightly different times. Consequently, its velocity information including speed and direction can be determined. The higher spatial resolution and more spectral bands of WorldView-2(WV2) imagery, compared to those of previous VHR satellites such as QuickBird and GeoEye-1, together with the new sensors' configuration of WV2, i.e., 4 bands on each side of the Pan sensor (MS1 and MS2), adds an opportunity to improve both moving vehicles extraction and the velocity estimation. In this paper, a novel processing framework is proposed for the automatic extraction of moving vehicles and determination of their velocities using single-pass WV2 imagery. The approach contains three major components: a) object-based road extraction, b) moving vehicle extraction from MS1 and MS2, and c) velocity estimation. The method was tested on two different areas of a WV2

[1] This paper has been published in *IEEE Journal of Selected Topic in Earth Observation and Remote Sensing:*
Salehi, B., Y. Zhang, M. Zhong, 2012. Automatic moving vehicle information extraction from single-pass WorldView-2 imagery, *IEEE Journal of Selected Topic in Earth Observation and Remote Sensing,* 5(1): 135-145.
For consistency

image, a high speed and a low speed traffic zone. Using a means of accuracy assessment, the method resulted in a "correctness" of 92% and a "completeness" of 77% for the extraction of moving vehicles. Furthermore, the estimated speeds and directions are very realistic and are consistent with the speed limits posted on the roads. The results demonstrate a promising potential for automatic and accurate traffic monitoring using a single image of WV2.

Keywords: WorldView-2, moving vehicle extraction, speed determination, time lag

5.1 INTRODUCTION

The high population growth and therefore high growth in road traffic volume over the last decades has increased road traffic to its congestion level [Campell, 2007]. This poses great challenges for today's road traffic research and planning. Vehicle monitoring is one of the important issues for modelling and planning of traffic and transportation management for both terrestrial and maritime areas. Traditionally, traffic monitoring has been conducted using ground-based equipment such as radar systems and video cameras [Munno *et al.,* 1993; Castellano *et al.,* 1999; Nag and Barnes, 2003]. These systems, however, have very narrow coverage area and are not able to observe global traffic situations. The use of remote sensing data for traffic monitoring has emerged as an alternative in recent years. Traffic monitoring using remote sensing data has the advantages of being fast and providing a wider coverage area over a long period of time. A number of studies have utilized high resolution aerial imagery (with resolution better than 0.3 m) [Ruskone *et al.,* 1996; Hinz and Baumgartner, 2001; Moon *et al.,* 2002; Toth and Grejner-Brzezinska, 2006; Hinz *et al.,* 2008; Yamazaki and Liu, 2008; Yao *et al.,*

2011] and LiDAR data [Toth and Grejner-Brzezinska, 2006; Yao *et al.,* 2011] for vehicle detection. Most of these studies are limited to vehicle detection and a very few attempts have been made to estimate the vehicle's velocity including speed and direction [Yamazaki and Liu, 2008; Yao *et al.,* 2011]. Aerial imagery and LiDAR data, however, are relatively expensive and not available for many areas.

Because of the sub-meter spatial resolution, very high spatial resolution (VHR) optical satellite images such as those of Ikonos (IK) and QuickBird (QB) have been utilized for vehicle detection, in recent years. Examples of vehicle detection using the Panchromatic (Pan) band of VHR imagery are [Sharma *et al.,* 2006; Zheng *et al.,* 2006; Jin and Davis, 2007; Zheng and Li, 2007; Eikvil *et al.,* 2009; Larsen *et al.,* 2009; Leitloff *et al.,* 2010; Larsen and Salberg, 2011]. Furthermore, some recent studies have estimated the velocity of vehicles using the small time lag between the Pan and multispectral (MS) sensors of QB [Leitloff *et al.,* 2010; Xiong and Zhang, 2008; Pesaresi *et al.,* 2008; Liu *et al.,* 2010].

The major challenge in moving vehicle information (position and velocity) extraction using VHR imagery is to extract the vehicle in two images (e.g., Pan and MS). Once the location of the vehicle is extracted the velocity is determined using the shift between two locations and the time lag between the two images. However, vehicle extraction using VHR imagery, particularly in MS bands, is a difficult task due to the relatively small size of vehicles. Most of the studies in this area have extracted the target vehicles either manually [Xiong and Zhang, 2008; Pesaresi *et al.,* 2008]or by incorporating ancillary data such as manually-digitized roads [Zheng *et al.,* 2006; Zheng and Li, 2007; Liu *et al.,* 2010] and existing road layers [Jin and Davis, 2007; Eikvil *et al.,* 2009; Larsen *et al.,*

2009; Leitloff *et al.,* 2010; Larsen and Salberg, 2011]. Furthermore, the aforementioned studies have detected vehicles in the Pan image and for those, in which the objective is the velocity estimation, the corresponding vehicle in MS image has been identified by utilizing a matching technique which is a computation-intensive process. Incorrect detection of the vehicle in MS image during the matching process is also a potential failure of matching methods. For example, [Leitloff *et al.,* 2010] observed a typical failure during the matching of Pan and MS images in their experimental results.

WorldView2 (WV2), launched October 2009, is the first VHR 8-band multispectral satellite which collects imagery with the highest spatial resolutions among all commercially-available VHR satellites. These unique spatial and spectral resolutions of WV2 along with its new sensors' configuration (Fig. 5.1) have added the opportunity for the improvement of moving vehicles information extraction. The WV2 satellite carries a Pan and two MS (MS1 and MS2) sensors onboard. The Pan sensor is located between the MS1 and MS2 and the sequence of collected images is MS1, Pan, and MS2 with approximately 0.13 seconds time lag between each MS and the Pan image and 0.26 seconds between the MS1 and MS2 [Smiley, 2011]. Consequently, a moving target is observed at three different positions by the satellite. The comparison of each pair of positions will result in velocity estimation. We believe that the velocity estimation by using the MS1 and MS2 images results in a higher accuracy than that by using the Pan and MS bands of QB imagery. This is because a) the MS1 and MS2 have the same spatial resolution and this ensures a more accurate comparison of the moving vehicle's positions in MS1 and MS2, which is a necessary step in velocity estimation, than that between the

Pan and MS bands of QB imagery, and b) the time lag between MS1 and MS2 of WV2 is about twice that between the Pan and MS bands of QB imagery which consequently results in more accurate velocity estimation. In addition, the new MS bands of WV2 (i.e., MS2 bands) have added the opportunity for road extraction, which is a necessary step in moving vehicle extraction, and also moving vehicle change extraction.

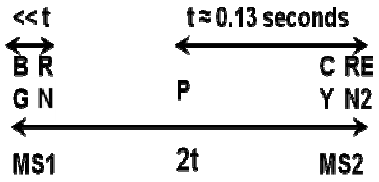


Figure 5.1 Time lags between WV2's sensors

In this study, we investigated a novel method for automatic moving vehicles' information extraction using WV2 imagery. In the proposed method, roads were first extracted using a hierarchical object-based image analysis framework of WV2 Pan-sharpened bands. Then, the changes in vehicles' locations in MS1 and MS2 images were enhanced (detected) using standardized principal component analysis (PCA) followed by an object-based change extraction (classification) of the areas within the boundary of roads (moving vehicles). Finally, the ground coordinates of vehicles' positions in MS1 and MS2 images were determined using the rational polynomial coefficients (RPCs) of the sensor followed by the calculation of vehicles' velocity using the change in the vehicle's positions and the time lag between MS1 and MS2 images.

The rest of the paper is structured as follows: Section 5.2 describes the methodology of this work which includes road extraction, moving vehicle detection, and velocity and direction determination. In section 5.3, the dataset is presented followed by the experimental results. Discussion of the results is presented in section 5.4. Finally, the conclusion of this work is presented in section 5.5.

5.2 METHODOLOGY

In this section, the method for moving vehicles information extraction is described. After a short overview, the specific parts of the extraction workflow are described in detail in the following subsections. Our proposed approach consists of three major parts: a) object-based road extraction (classification), b) moving vehicle extraction using PCA-based change detection of MS1 and MS2 images followed by object-based extraction of changes within the boundary of roads and c) Determining the ground positions of the vehicle in MS1 and MS2 images using the sensor model parameters, provided by the satellite, and finally, estimating the velocity of the vehicle using the time lag between two MS images. The flowcharts of the first two components are illustrated in Fig. 5.2 while that of the third component is shown in Fig. 5.3.

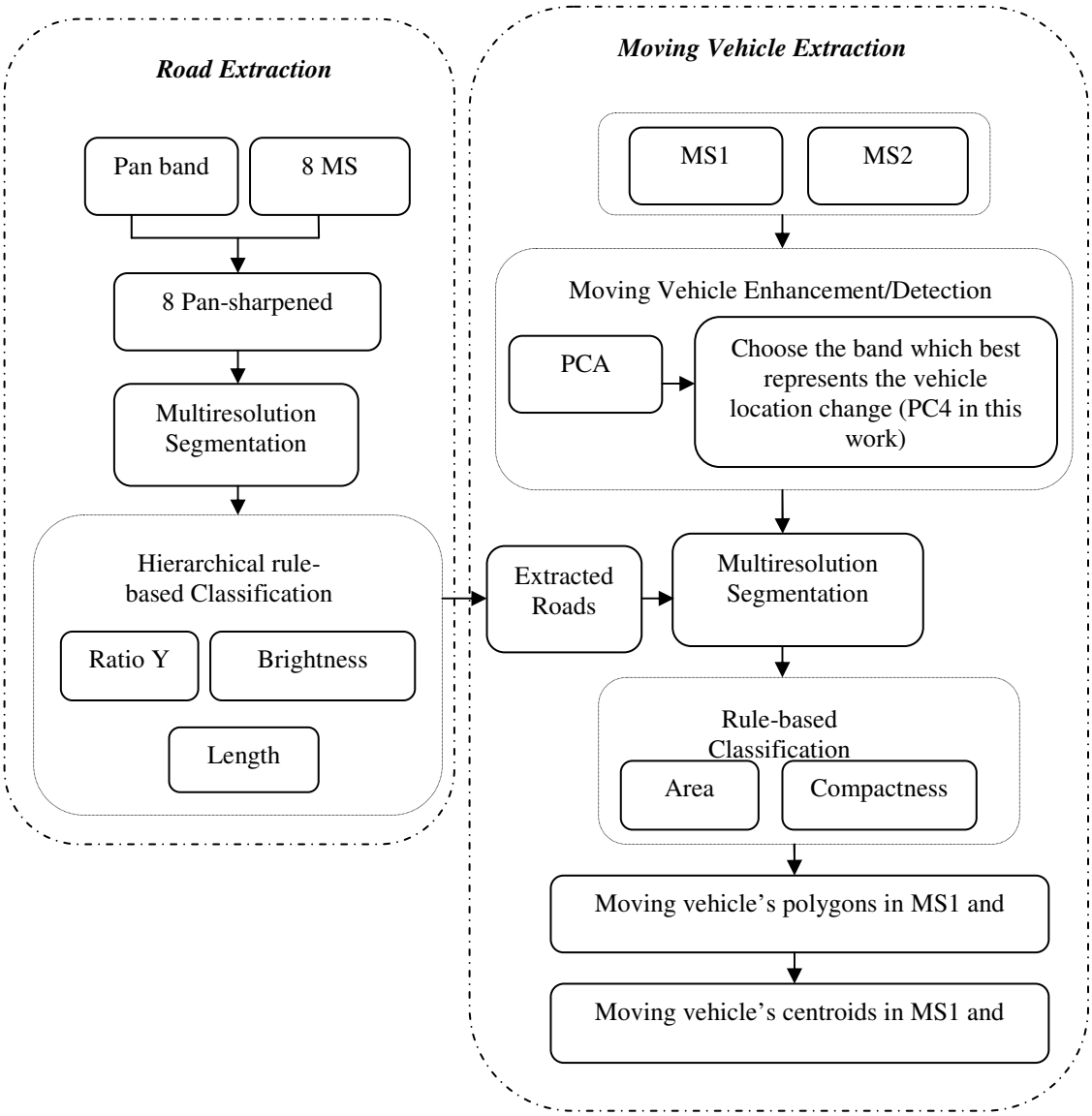


Figure 5.2 Flowchart of the proposed object-based moving vehicle extraction

5.2.1 Road Extraction

Vehicles move along roads and, thus, the search area for moving vehicle detection can be restricted to roads. However, road extraction from VHR imagery is not an easy task due to the spectral similarity of roads to parking lots and buildings. Most of the studies for

vehicle extraction using VHR imagery have extracted roads either manually [Zheng *et al.,* 2006; Zheng and Li, 2007; Liu *et al.,* 2010] or by using archived GIS road layers [Jin and Davis, 2007;; Eikvil *et al.,* 2009; Leitloff *et al.,* 2010; Larsen and Salberg, 2011]. In this work, however, we developed a semi-automatic object-based road extraction framework using the Cognition Network Language available in eCognition® software package. The flowchart of object-based road extraction is illustrated in the left part of Fig. 5.2. The process starts with Pan-sharpening of the image followed by segmenting the Pan-sharpened image and the process ends by classifying the segmented image into roads and non-roads. The following discusses each step in the flowchart in detail.

In order to take advantage of both the spectral information of eight MS bands and the high spatial resolution of the Pan band in the road extraction procedure, the Pan and eight MS bands of the WV2 image were fused using the UNB Pansharp algorithm [Zhang, 2004] resulting in eight Pan-sharpened bands.

The first step in object-based image analysis is segmentation, which is the process of partitioning the image into a set of discrete non-overlapping regions (image objects) on the basis of internal homogeneity criteria [Devereux *et al.,* 2004] . In the second step, image objects are assigned to a class based on the fuzzy membership function of objects' features such as spectral (e.g., mean brightness value, ratio of bands) and morphological (shape, size, etc) features.

For road extraction, due to the presence of roads with different widths (from wide highways to narrow city streets) in typical urban environments, the multiresolution segmentation is preferred. The multiresolution segmentation is a bottom up procedure

based on the Fractal Net Evolution Approach (FNEA) [Baatz and Schape, 2000], which is embedded in eCognition software package. It starts from pixel level and grows up based on a pair-wise region merging technique [eCognition, 2010a]. The results are objects with different sizes and shapes in different levels. Lower levels represent smaller objects while higher levels represent larger objects. Three key parameters, namely shape, compactness, and scale need to be set up in multiresolution segmentation. Usually the last two parameters, which control the homogeneity of objects, are set as eCognition default. Scale, which is considered as the most crucial parameter of multiresolution segmentation [Myint *et al.,* 2011], controls the objects' size. Higher values for scale parameter result in larger image objects (higher levels) and lower values in smaller image objects. The details of the steps involved in multiresolution segmentation can be found in [eCognition, 2010b].

Following the segmentation, a hierarchical rule-based classification schema was developed to classify the road areas. The rule-based approach allows the analyst to combine spectral and morphological features of objects to assign a class membership degree (between 0 and 1) to each object based on a fuzzy membership function. Furthermore, it has hierarchical capability meaning that the approach begins to classify the entire image into general classes (e.g., traffic areas and non-traffic areas) and then, by utilizing other features, each class is split to more detailed subclasses (e.g., roads and parking lots)

In this work, the image was segmented into four levels and objects in the fourth level were utilized in the rule-set for the subsequent classification. We experienced the Ratio Y

and the Brightness value in band NIR2 of objects are two very helpful spectral features in separating traffic areas (including roads and parking lots) from other areas in the image [Salehi *et al.,* 2011]. The Ratio Y describes the amount that band Y contributes to the total brightness for an object and is calculated as the ratio between the brightness value of band Y over the summation of brightness values of all bands [eCognition, 2010b]. The last step in the rule-set hierarchy is to separate roads from parking lots. Roads are elongated objects and they can be distinguished from parking lots by a Length threshold. Length has high value for road and low value for parking lot objects.

5.2.2 Moving Vehicle Extraction

To determine the velocity of a vehicle, its relative displacement in two images needs to be determined. The conventional method is to extract the vehicle in one image (i.e., the PAN image) and then to determine the corresponding vehicle in the other image (i.e., the MS image) using a matching technique and then comparing the positions of the vehicle in each image. However, in this article, to restrict the vehicle extraction to the moving ones that are of interest (not the stationary ones), a different strategy was employed. The strategy begins with detecting (enhancing) the change areas between the MS1 and MS2 images, which are moving vehicles, followed by extracting the change areas.

A moving vehicle has different positions in MS1 and MS2 images. Consequently, if a change detection procedure is applied to the stacked set of the MS1 and MS2, the change areas will represent moving vehicles. Indeed, the result of change detection for a moving vehicle will be two neighbor objects which each represent the vehicle in each image (Fig. 5.6, PC4). These neighbor objects are then extracted using the classification of the change

image. The flowchart of the proposed moving vehicle extraction is presented in the right part of Fig. 5.2. The flowchart includes two steps: a) moving vehicle detection (enhancement) by applying change detection to all eight MS bands; and b) extracting the changes (moving vehicles) using the object-based classification. Principal Component Analysis (PCA) was employed for change detection because of its simplicity and capability of enhancing the information on change [Deng *et al.,* 2008].

5.2.2.2 PCA-based Change Detection of Moving Vehicles

Here, we are encountering a change detection problem in which the targets of interest are moving vehicles and the multi-temporal images are the MS1 and MS2 images. The PCA is a commonly used statistical method for many aspects of remote sensing including change detection. Basically, PCA is an orthogonal transformation which converts a set of possibly correlated variables (image bands) into a set of uncorrelated bands called principal components (PCs). A new set of coordinate axes is calculated from the eigenvectors which in the unstandardized form are calculated from the covariance matrix, and in standardized form are calculated from the correlation matrix. The first new axis or principal component is oriented in the direction of maximum variance, with subsequent axes sequentially rotated in the direction of the largest remaining variance. Studies indicate that the standardized PCA appears to be more effective than the unstandardized one [Deng *et al.,* 2008]. For this reason, and also based on our experimental results, standardized PCA was utilized in this work. While the first PCs, corresponding to large eigenvalues, reflect the unchanged part of the image, the last PCs reflect the changed part. The difficulty in PCA-based change detection, however, is automatically

determining which PC represents change areas [Radke *et al.,* 2005]. In the present study, we performed this task by visual inspection (see Fig. 5.6). We refer to this PC as PC change.

5.2.2.3 Object-based Moving Vehicle Extraction

Having enhanced the changes in moving vehicles by the PCA, the change areas will be extracted from the PC change image. Similar to the road extraction method, an object-based image analysis framework was developed to extract the enhanced moving vehicles from the PC change image. As the roads were already extracted, moving vehicle extraction is restricted to the road areas of PC change image.

Similar to road extraction, the multiresolution segmentation was utilized for vehicle extraction. In contrast to the road extraction step, in which the entire Pan-sharpened image was segmented, for moving vehicle extraction only the road areas of PC change were segmented into three levels. Level one represents small objects (smaller than a normal vehicle size) while objects in level three are more meaningful [Tian and Chen, 2007] for representing vehicles size and shape (see Fig.5.7). In the PC change image, for each moving vehicle there are two neighboring objects with very dark and very bright gray values. Each object represents the vehicle in one of the MS images (MS1 or MS2). Therefore, if two such neighboring objects are extracted, the center of each object represents the centroid of the moving vehicle. Having tested several spectral and morphological features of the objects in third level (Fig. 5.6c), we learned that very dark and very bright objects are well extracted using their morphological features including shape and area. In other words, vehicle' objects possesses a specific shape and area

(number of pixels) and thus are distinguishable from the rest of the road by applying certain thresholds to the shape and area of all objects. *Compactness*, as a shape feature, is defined as the area of smallest rectangle enclosing the image object (bounding box) divided by the total area of the object [eCognition, 2010b], was used in this study.

5.2.3 Determining the Ground Positions and Velocity of Moving Vehicles

From the previous section, the centroids (image position) of moving vehicles in both MS1 and MS2 images were extracted. The next step is to determine the respective positions of those centroids in ground space.

Fortunately, VHR satellites such as QB and WV2 provide a set of sensor model parameters or rational polynomial coefficients (RPCs), which describe the image-to-object space transformation. The transformation can be given as:

$$\begin{cases} x = \frac{P_1(X,Y,Z)}{P_2(X,Y,Z)} \\ y = \frac{P_3(X,Y,Z)}{P_4(X,Y,Z)} \end{cases} \quad (1)$$

$$P_1(X,Y,Z) = a_1 + a_2Y + a_3X + a_4Z + a_5YX + a_6YZ + a_7XZ + a_8Y^2 + a_9X^2 + a_{10}Z^2 + a_{11}XYZ + a_{12}Y^3 + a_{13}YX^2 + a_{14}YZ^2 + a_{15}Y^2X + a_{16}X^3 + a_{17}XZ^2 + a_{18}Y^2Z + a_{19}X^2Z + a_{20}Z^3$$

$$P_2(X,Y,Z) = b_1 + b_2Y + b_3X + b_4Z + \cdots b_{20}Z^3$$

$$P_3(X,Y,Z) = c_1 + c_2Y + c_3X + c_4Z + \cdots c_{20}Z^3$$

$$P_4(X,Y,Z) = d_1 + d_2Y + d_3X + d_4Z + \cdots d_{20}Z^3$$

In the above equations, x, y are the normalized (offset and scaled) image coordinates (row, column) and X, Y, Z are the corresponding ground coordinates. $a_i, b_i, c_i, and\ d_i$ are coefficients which are supplied with WV2 image products.

According to WV2 sensor specifications [DigitalGlobe, 2009], the positional accuracy of well-identified objects in the image (circular error with the confidence level of 90%) of the sensor model is in the range of 4.6 to 10.7 m (excluding terrain and off-nadir effects) without using ground control points (GCPs), and 2.0 m with using GCPs. Obviously, this level of positional accuracy does not meet the requirements for speed determination (as it produces an error of speed determination that may range 63 km/h to 148 km/h). The objective of this research is to determine the moving vehicles' velocity. Therefore, relative positions of the vehicles in MS1 and MS2 are of major interest and the above mentioned absolute geo-positioning error can be neglected if the topography around the moving vehicles is changing smoothly. This is a feasible condition for vehicles as the road networks in general do not show very steep height gradients [Pesaresi *et al.,* 2008].

Fig. 5.3 illustrates the steps involved in the moving vehicles' information extraction workflow. Having determined the ground positions of moving vehicles' centroids in MS1 and MS2 using RPCs, the displacement vector of changes in vehicle's position is

determined. This vector represents the distance between two positions of the vehicle in MS1 and MS2 as well as its moving direction. Consequently, using this distance and the time lag between MS1 and MS2 (0.26 s), the vehicles' speeds are calculated.

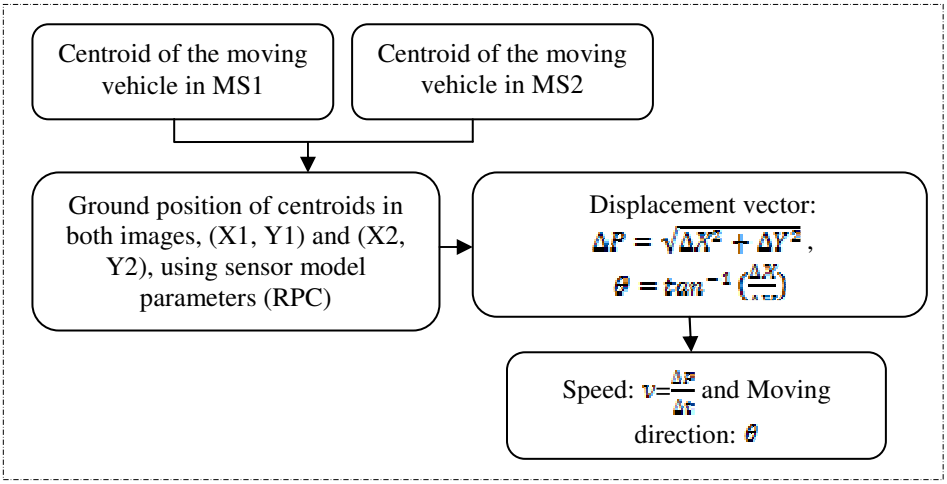


Figure 5.3 Flowchart of the moving vehicles information (position, speed and direction) extraction.

5.3 EXPERIMENTAL RESULTS AND DISCUSSION

5.3.1 Dataset

The test area is a part of the City of Moncton in New Brunswick, Canada. A subset of a geometrically corrected WV2 image, acquired on October 5, 2010, was used in this study. The WV2 imagery includes a Pan band and two sets of multispectral bands (MS1 and MS2). The spatial resolution of the Pan and MS bands are 0.46 m and 1.84 m at nadir, respectively. However, the image used in this study has the resolutions of 0.5 m and 2 m in the Pan and MS modes. The MS1 contains the conventional multispectral bands (i.e., blue: B, green: G, red: R, and near infrared: NIR1), and the MS2 possesses four newly added bands which are coastal (C), yellow (Y), red edge (RE) and near infrared-2 (NIR2) [34]. The sequence of images collected by WV2 is MS1, Pan and MS2.

The time lag between MS1 and MS2 is about 0.26 seconds [Smiley, 2011]. This feature of the satellite is the basis for extracting the moving vehicle information from single-pass WV2 imagery. The image contains different traffic areas including a highway, a large traffic circle, a road intersection, and many parking lots. This variety of traffic areas is good for testing our proposed approach for different traffic conditions. Fig. 5.4 shows a true color composite of the Pan-sharpened WV2 image used in this study. Rectangles A and B represent two different areas in terms of vehicle velocity (i.e., high speed (A), and moderate speed (B) zones). These two areas will later be used to show the final results of the proposed method.

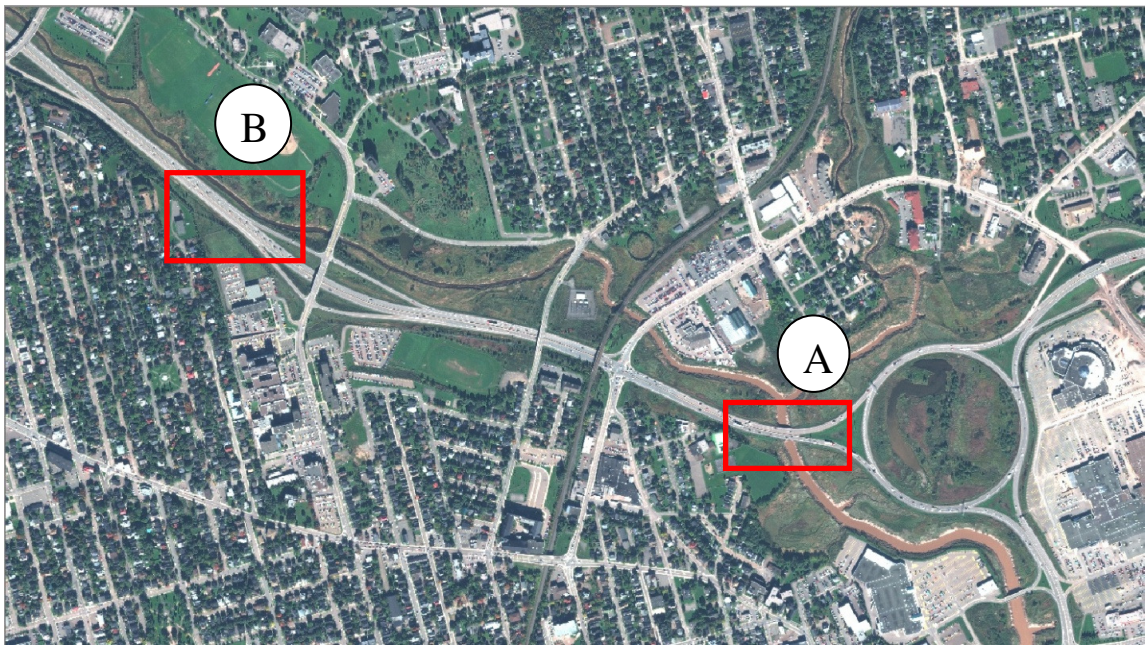


Figure 5.4 The true color composite of the Pan-sharpened WV2 image of the Test area. Rectangles A and B represent the areas which will be used later to show the final result of the proposed method.

5.3.2 Road Extraction Results

The object-based road extraction results can be seen in Fig. 5.5. As it is observed, the entire highway, the roundabout, and most of the minor roads/streets were extracted. Two new spectral bands of WV2 (Y and NIR2) make the major contribution in distinguishing traffic areas (roads and parking lots) from the rest of the image. Moreover, the capability of object-based classification in utilizing spatial features as well as the spectral features of objects, resulting from the segmentation step, helps the classifier to separate roads and parking lots. As mentioned, in the previous section, we utilized the length of objects as a shape feature to separate roads from parking lots. Roads are elongated features and can be separated from parking lots if a proper length threshold is set up in the classification rule-set. Although commission error exists in the classification of roads (some parking lots are misclassified as roads), this will not affect the final result of the moving vehicle extraction. This is because the vehicles in parking lots are mostly stationary and there is no change in their positions in two images. Consequently, the PCA change detection algorithm, in the next step, will not detect them as moving vehicles.

Figure 5.5 The result of the proposed object-based road extraction overlaid on the Pan-sharpened image. The extracted roads are the blue areas.

5.3.3 Moving Vehicle Extraction Results

The PCA technique for change detection requires the separate images first be stacked in a multi-temporal composite image [Sunar, 1998]. In this work, the MS1 and MS2 of the WV2 are considered as two images. Since WV2 MS bands have a spatial resolution of 2 meters, small vehicles are not represented clearly in the image. Subsequently, the PCA does not yield a clear change of the moving vehicle. Therefore, MS1 and MS2 bands were first resampled to the spatial resolution of the Pan band (0.5 m) using the cubic convolution resampling and then the PCA was applied to the resampled eight multispectral bands. The first five components of PCA are depicted in Fig. 5.6. Although the first component (PC1) represents the major part of the information, the changes (i.e., moving vehicles) are well detected in PC4 (PC change). The moving cars in the highway

are detected as two neighboring black and white objects, while surprisingly the queue of the stationary cars (in front of stoplights) in the road intersection on the top-left corner of the PC4 is not detected as moving cars. These cars are represented as a row of black objects in the top arm of intersection (see also Fig. 5.8 (a)).

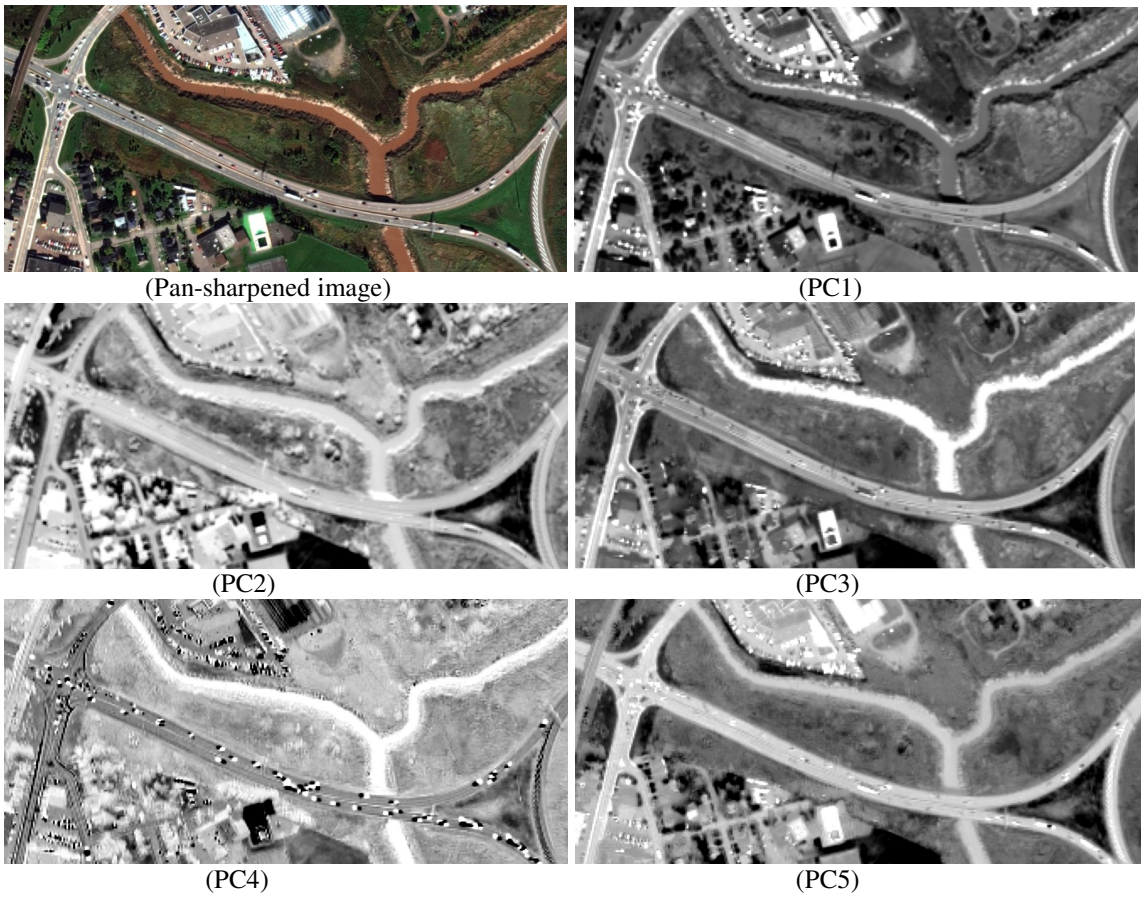
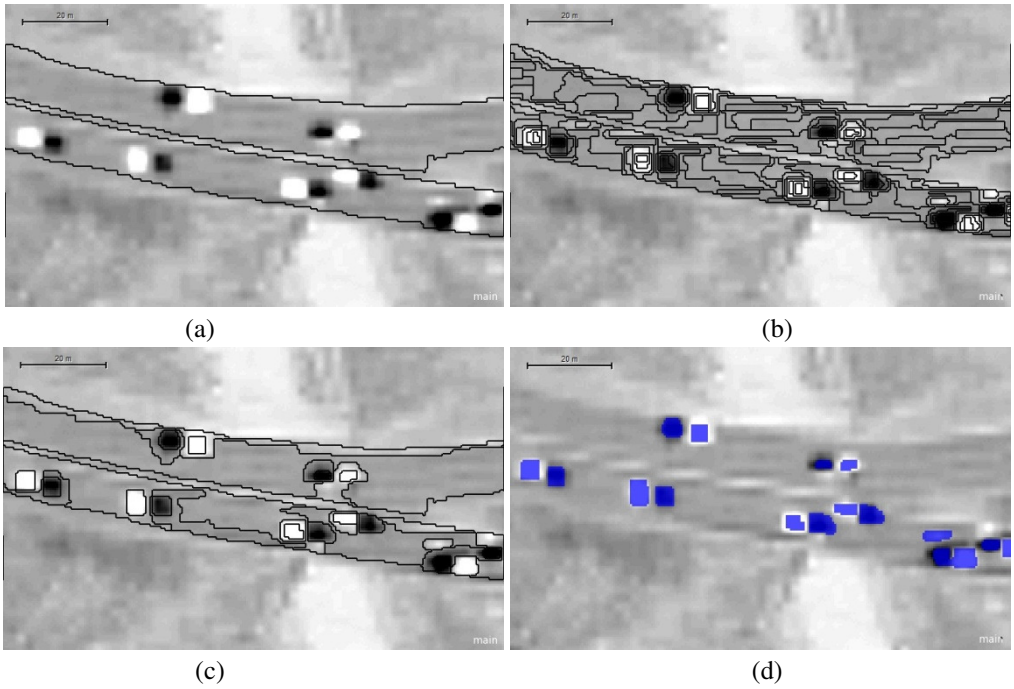


Figure 5.6 The Pan-sharpened image along with the first five components of PCA applied to the resampled 8 multispectral bands of WV2. Moving vehicles can be clearly seen in PC4 (PC change).

Fig. 5.7 shows the first and third levels of segmentation as well as the classification result for a portion of a road in the image. Segments in the third level are more meaningful for representing

the vehicles and the classification result shows that all vehicles in this sample image were extracted.



(a) (b)

(c) (d)

Figure 5.7 The segmentation and classification results of a portion of a road in PC4 image .(a) Extracted road from the previous step, (b) first, and (c) third level of road segmentation and (d) classified (extracted) moving vehicles. The two neighboring objects in (d) represent the vehicle's location in MS1 and MS2 images.

The final results of our moving vehicle extraction for two larger subsets of the study area are depicted in Fig. 5.8. The extracted moving vehicles are red objects overlaid on the Pan-sharpened image of WV2. As can be seen, most of the moving cars have been extracted in both images. Interesting results can be seen in this Figure where the stopped cars in the roads' intersection (top-left corner of Fig. 5.8 (a)) and those parked in the parking lot have not been extracted as moving vehicles, since the PCA had not detected them in the previous step. In Fig. 5.8 (b), almost all cars in motion have been extracted by

the classifier. Furthermore, the cars parked in the parking lots located in the bottom-middle of this figure were not extracted as moving cars. An enlarged version of the extracted vehicles is depicted in Fig. 5.10. In this figure, the centroids of each pair of neighboring polygons represent the location of the vehicle in the MS1 and MS2 images.

(a)

(b)

Figure 5.8 The results of the proposed moving vehicle extraction approach overlaid on the Pan-sharpened image for two different parts of the study area. (a) low speed zone, and (b) high speed zone. The extracted vehicles are red objects.

In order to validate the moving vehicle extraction results, a numerical accuracy assessment was conducted by comparing the number of vehicles found by manual inspection to that found by our proposed approach. The following three categories can be defined by comparing the automatically extracted and manually found vehicles. (See also Fig. 5.9).

True Positive (TP): the number of correctly extracted moving vehicles

False Positive (FP): the number of incorrectly extracted moving vehicles

False Negative (FN): the number of missing moving vehicles

In these categories, vehicles are counted in both MS1 and MS2 images, since moving vehicles have two different positions in MS1 and MS2. For instance, if both positions of a moving vehicle (in MS1 and MS2) are correctly extracted, the TP value for the vehicle is two. Similarly, if only one position of the vehicle is extracted, the TP and FN values are both one. Fig. 5.9 illustrates the three categories defined for a small part of a road in the study area. Based on these three categories the following statistical measures are computed [Agouris *et al.,* 2004; Nie, 2006].

$$Correctness = \frac{TP}{TP+FP} \qquad (2)$$

$$Completness = \frac{TP}{TP+FN} \qquad (3)$$

$$Quality = \frac{TP}{TP+FP+FN} \qquad (4)$$

Correctness and completeness are the converse of commission and omission errors, respectively. These two measures are complementary and need to be interpreted

simultaneously. Quality is the normalized measures between correctness and completeness, which shows the overall accuracy of the extraction approach. Table 5.1 reports the three calculated measures for the two different parts of the study area shown in Fig. 5.8.

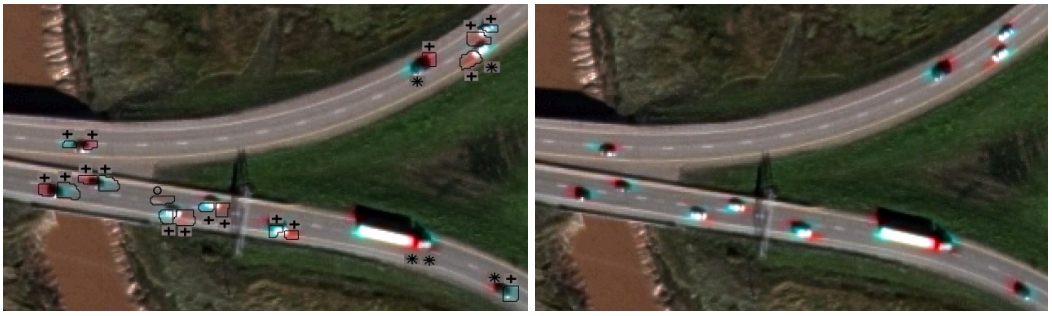


Figure 5.9 Moving vehicles and their extracted polygons: (cross) True positive, (circle) false positive, and (asterisk) false negative

Table 5.1 Accuracy assessment for the moving vehicle extraction: Results were reported for low speed and high speed zones shown in Fig. 5.8

Area	TP	FP	FN	Correctness	Completeness	Quality
Fig.5.8 (a)	114	11	34	91.2%	77.0%	71.7%
Fig.5.8 (b)	96	8	28	92.3%	77.4%	72.7%
Total	210	19	62	91.7%	77.2%	72.2%

Most of the studies in the field of velocity estimation using VHR imagery (i.e., using the time lag between the PAN and MS bands) have reported the vehicle extraction results from only the PAN band but not from the MS bands. For example, [18] reported the correctness, completeness, and quality measures of 95.3%, 82.5%, and 79.2%, respectively for the vehicle extraction in the PAN band of QB. Another example is [22]

where the authors obtained the user accuracy (correctness) and producer's accuracy (completeness) of 89% and 94% using the Pan band of QB [Liu *et al.,* 2010]. These accuracy measures are slightly higher than those achieved in our work (Table 5.1). However, as mentioned, they were achieved only on the Pan band but not on the MS bands. For the velocity estimation, the vehicle in both Pan and MS imagery is required. Vehicle extraction from the MS bands leads to worse results, due to the coarser resolution of MS bands compared to the PAN band. The accuracies achieved in this article (Table 5.1), however, show the vehicle extraction results in both MS1 and MS2 images. Furthermore, these studies have used the GIS road layer [Leitloff *et al.,* 2010] or manually extracted roads [Liu *et al.,* 2010] in their vehicle extraction algorithms, while in this article road extraction is a part of the proposed method.

5.3.4 Moving Vehicles Information Extraction Results

Moving speed and direction of the vehicles for two different subsets of WV2 imagery were determined using the approach described in Fig. 5.3. We tried to pick two different areas (in terms of vehicles' speed). These two areas (Area A and Area B) are shown with red rectangles in Fig. 5.4. The enlarged versions of Area A and Area B are depicted in Fig. 5.10. Vehicles in Area A are either approaching (vehicles 1-8) or just have passed the traffic circle (vehicles 9-13) and thus should have relatively low speeds. Vehicles in Area B, however, are running on the highway and thus should have higher speeds compared with vehicles in Area A. Using the "Street View" capability of Google Maps

(http://maps.google.ca), we checked the posted speed limit for routes shown in Fig. 5.10 and they were 60 km/h and 100km/h for route A and route B, respectively.

The calculated moving speed and direction of the extracted vehicles shown in Fig. 5.10 are reported in Table 5.2 and Table 5.3. As expected, vehicles in Area A have relatively low speeds ranging from 46 km/h to 100 km/h except vehicle 13 which is moving at 119 km/h. Although the calculated speeds are relatively higher than the posted speed limit, the results represent realistic speeds, as the image has been acquired around 3:40 pm when the traffic flow is light and drivers run faster. Vehicle 13 is running almost twice the speed limit! The moving azimuth for this vehicle is about 303 degrees which is larger than that of the vehicle behind (vehicle 12). It shows that vehicle 13 just over took vehicle 12 and is moving from the high speed lane to the low speed lane (see also Fig. 5.10 A) and therefore should have higher speed compared to vehicle 12.

As can be observed in Table 5.3, vehicles in Area B have relatively high speed ranging from 72 km/h to 128 km/h. These speeds are also realistic as the vehicles are running on a faster highway as compared to those in Area A. An interesting result in Table 5.3 is that all vehicles moving east to west (vehicles 16-26) have speeds below 110 km/h, while vehicles 14 and 15 are running at 122 and 128 km/h, respectively. The reason is that vehicles 16-26 are coming from a low speed zone and also the traffic is heavy on this route while for vehicles 14 and 15 the traffic is light enough to run at a high speed.

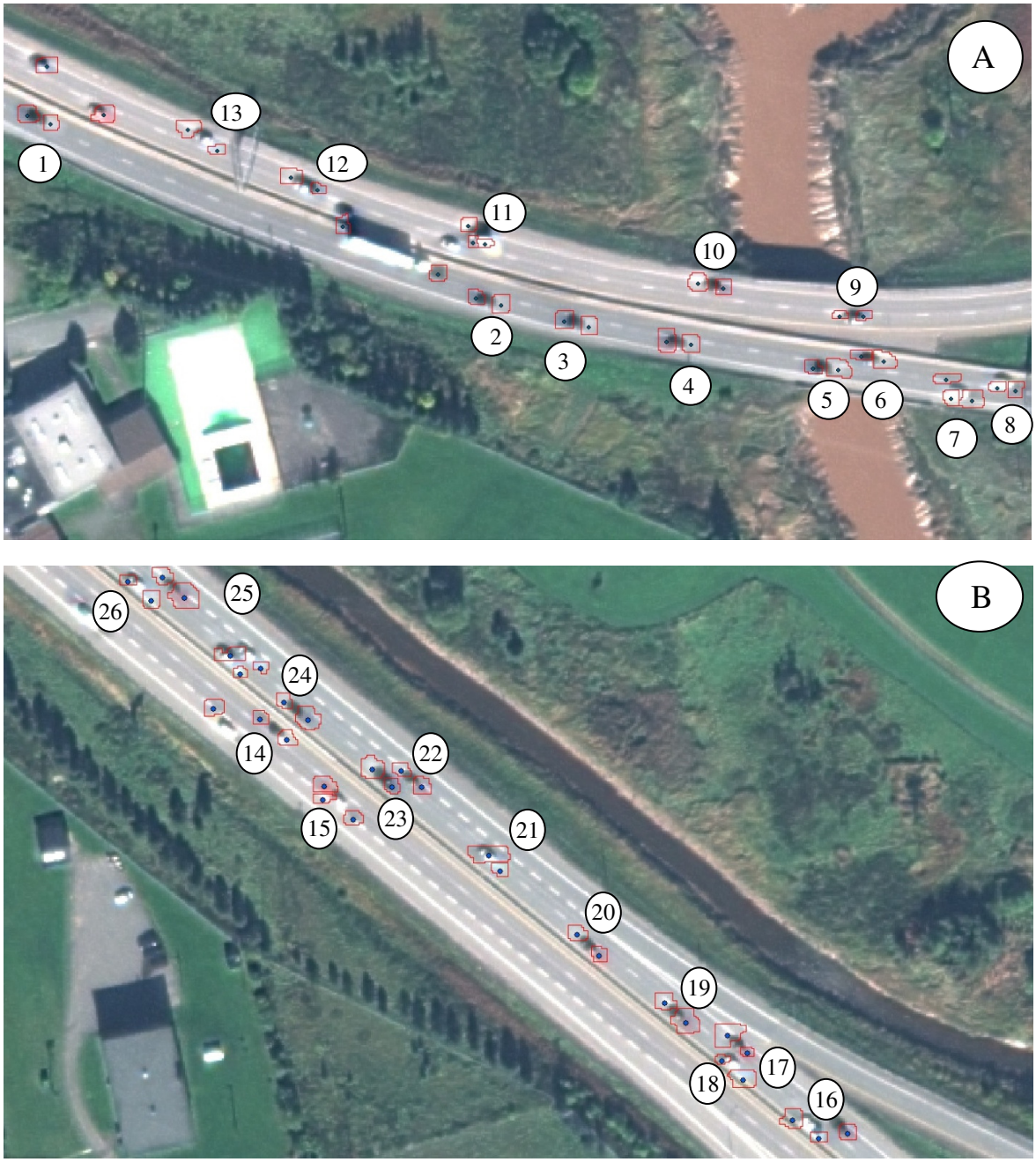


Figure 5.10 The extracted polygons and centroid of moving vehicles' position in MS1 and MS2 overlaid on the Pan-sharpened image for two different parts of the study areas (A is a relatively low speed and B is relatively high speed route).

Table 5.2 The positions (in MS1 and MS2), speed and moving direction of moving vehicles shown in Fig. 5.10 (A), determined using the proposed approach

	MS1		MS2		Speed (km/h)	Moving Azimuth(º)
Vehicle #	X(m)	Y(m)	X(m)	Y(m)		
1	362,846.5	5,106,437.1	362,852.6	5,106,434.9	90	109.9
2	362,959.7	5,106,390.7	362,966.1	5,106,389.2	92	102.8
3	362,982.4	5,106,385.0	362,988.5	5,106,383.7	87	101.9
4	363,008.1	5,106,380.3	363,014.4	5,106,379.4	88	98.3
5	363,044.8	5,106,373.5	363,051.6	5,106,373.1	95	93.1
6	363,057.5	5,106,376.1	363,063.4	5,106,375.0	83	100.6
7	363,080.3	5,106,365.5	363,085.3	5,106,365.4	69	92.1
8	363,091.9	5,106,368.3	363,096.4	5,106,367.6	62	99.4
9	363,058.1	5,106,386.4	363,051.8	5,106,386.4	87	270.0
10	363,022.4	5,106,394.0	363,016.1	5,106,394.8	88	276.7
11	362,962.6	5,106,404.6	362,959.3	5,106,405.1	46	279.4
12	362,919.5	5,106,418.6	362,912.9	5,106,421.6	100	293.9
13	362,894.6	5,106,428.6	362,887.4	5,106,433.3	119	303.1

Table 5.3 The positions (in MS1 and MS2), speed and moving direction of moving vehicles shown in Fig. 5.10 (B), determined using the proposed approach

	MS1		MS2		Speed (km/h)	Moving Azimuth(º)
Vehicle No.	X(m)	Y(m)	X(m)	Y(m)		
14	361,686.2	5,106,950.6	361,693.1	5,106,945.1	122	128.2
15	361,702.6	5,106,929.6	361,710.5	5,106,924.7	128	122.1
16	361,830.6	5,106,842.4	361,823.8	5,106,847.0	113	304.5
17	361,812.1	5,106,864.3	361,807.1	5,106,868.9	92	312.5
18	361,811.1	5,106,857.3	361,805.9	5,106,862.4	102	314.5
19	361,796.3	5,106,872.2	361,791.0	5,106,877.4	102	314.4
20	361,773.9	5,106,889.5	361,768.3	5,106,895.0	110	315.0
21	361,748.6	5,106,911.3	361,745.4	5,106,915.4	72	322.4
22	361,728.2	5,106,933.1	361,722.9	5,106,937.2	92	307.9
23	361,720.6	5,106,933.2	361,715.2	5,106,937.8	98	310.7
24	361,698.6	5,106,950.4	361,692.5	5,106,954.8	104	313.8
25	361,666.9	5,106,982.1	361,661.3	5,106,987.4	107	313.8
26	361,658.3	5,106,981.4	361,652.0	5,106,986.3	110	307.5

5.4 DISCUSSION

This study demonstrates the very high potential of WV2 imagery when used in conjunction with object-based image analysis for the extraction and velocity estimation of moving vehicles. The very high spatial resolution, along with eight multispectral bands and the relatively large time lag between the MS1 and MS2 bands of WV2 imagery (compared to that between the PAN and MS bands of QB) have increased the opportunity for accurate extraction and velocity estimation of the moving vehicles. Furthermore, the ability of object-based image classification in employing both spectral and spatial information (including shape, texture, and context) of the image has increased the accuracy of extracting the roads and moving vehicles.

The proposed object-based road extraction shows success with the extraction of major roads and most of the minor roads (Fig. 5.5). The presented hierarchical rule-based method begins to classify the image as traffic (including parking areas and roads) and non-traffic areas. Because of their spectral similarity, the separation of built-up from traffic areas using the conventional bands of VHR imagery is a challenging task. However, the new bands available in WV2, particularly bands Y and NIR2, proves a great aid to distinguishing traffic areas from the rest of image including built-up areas. The next step in the object-based rule-set hierarchy is to separate roads from parking areas. This is also challenging because these two classes are very similar in terms of their spectral values. Unlike pixel-based classification approaches, which rely mainly on the spectral information of individual pixels, in object-based approaches, both the spectral and spatial information of objects (e.g., shape and size), resulting from the segmentation

step, are utilized. The use of shape measures, including the length of objects, is particularly effective for the separation of roads from parking lots due to the distinguished morphological properties of roads (i.e., roads are elongated features).

Fig. 5.5 shows some minor roads that have not been completely extracted. The main reason is that these roads are partially covered by tree branches and thus they have been missed during the classification. Also, as mentioned in the result section, some parking lots have been misclassified as "roads" in Fig. 5.5. This will not affect the final moving vehicle extraction results, as the vehicles in parking lots are stationary and thus the change detection algorithm in the next step will not detect them.

Applying the PCA to the stacked set of the MS1 and MS2 images provides a very satisfactory result in enghancing the changes occuring on the roads (i.e., the moving vehicles). The object-based moving vehicle extraction method uses the PC change image as the input. In the PC change image, a moving vehicle is presented as two nighboring objects with bright and dark brightness values (Fig. 5.7a). As a consequence, the contrast of the vehicle to background road is enhanced. Morover, the low contrast caused by the colour of vehicles is avoided. Consequently, the segmentation and the following classification of moving vehicles in the PC change image is more reliable than in the original MS images.

The proposed moving vehicle exctaction method resulted in a correctness and completeness of about 92% and 77%, respectively. Some vehicles were not extracted (false negative), which causes the relative low completeness value. The missed vehicles are mainly large size vehicles (e.g., trucks). This is mainly because the classification

decision rule is based on the shape and size of objects, and any object greater than the threshold remains unclassified. This mis-classification can be reduced by the incorporation of other spatial features of objects such as textural feature into the classification rule, particularly for traffic areas contaning moving vehicles with different sizes.

Considering the moving directions and traffic conditions, the estimated velocities are realistic and are consistent with the speed limits posted on the roads. The accuracy of speed estimation largely depends on the determined centriod positions, which, in turn, rely on the performance of segmentation. More meaningful segments for vehicles (i.e., segments that are closer to the real outline of vehicles) result in a more accurate estimation of centroids and, consequently, a better estimation of speeds and moving directions. The focus of our future work will be to improve the segmentation part of the method. Another factor that affects the estimation of speeds is the time lag between the two images. A larger time lag would result in a better estimation of speeds, as speed estimation is less affected by the possible error in the positions of the vehicle . The time lag between MS1 and MS2 images is about twice that between the PAN and MS bands of QB imagery and,therefore, it leads to a more accurate estimation of speeds.

5.5 CONCLUSION

A new method for automatic moving vehicle information (position and velocity) extraction using WV2 imagery was presented. The method starts with a hierarchical rule-based object-based road extraction (classification) using the Pan-sharpened bands of the WV2 imagery. Then, a standardized PCA is applied to the MS bands to enhance (detect)

the changes in vehicle locations in MS1 and MS2 images. A rule-based object-based classification is then developed to extract the enhanced changes within the boundary of roads (moving vehicles). The result of this step is individual polygons representing vehicle locations and their corresponding centroids in MS1 and MS2 images. Finally, the ground coordinates of the centroids are determined using the image-to- ground space transformation given by the sensor's RPCs, followed by calculating the moving direction and speed using the shift in vehicle positions and the time lag between the MS1 and MS2 images. The experimental result shows that the proposed method can effectively extract roads and moving vehicles within the boundary of roads and estimate their velocities and moving directions. Further, the method is automatic and can be transferred to WV2 imagery of other areas. This promises a very high potential of WV2 imagery in automatic and cost effective monitoring of moving targets.

We recognize that there is still potential for improvement in the object-based extraction of enhanced changes (moving vehicle). Any improvement in segmentation (to get more meaningful objects for vehicles) and classification of PC change image can increase the accuracy of vehicle's centroid extraction. This will be the focus of our future research.

ACKNOWLEDGEMENTS

The WorldView-2 image was provided by DigitalGlobe® Inc. to the authors through "The DigitalGlobe 8-Band Research Challenge" contest. The authors are grateful to DigitalGlobe for making such a high quality dataset available. This research is supported by the NSERC (Natural Science and Engineering Research Council of Canada) Strategic

Project Grants and NSERC Discovery Grants Programs. The authors would like to thank these two agencies.

REFERENCES

Agouris, P. ,P. Doucette, and A. Stefandis, "Automation and digital photogrammetric workstations," In: McGlone, J.C., Mikhail, E.M., Bethel, J., Mullen, R. (Eds, Manual of Photogrammetry, fifth ed. American Society for Photogrammetry and Remote Sensing, Bethesda, MA, pp. 949-981, 2004.

Baatz, M. , and A. Schape, "Multiresolution segmentation—an optimization approach for high quality multi-scale image segmentation," *In Proc*. Angewandte Geographische Informationsverarbeitung XII, Beitra¨ge zum AGIT-Symposium Salzburg, 2000.

Campell, G. (2007). Office for Official Publications of the European Communities, *Panorama of Transport.* [Online]. Available: http:// epp.eurostat.ec.europa.eu/ cache/ITY_OFFPUB/ KS-DA-07-001/EN/KSDA- 07-001-EN.PDF.

Castellano, G. , J. Boyce, and M. Sandler, "CDWT optical flow applied to moving target detection," *IEE Colloquium on Motion Analysis and Tracking*, pp.17/1-17/6, 1999.

Deng, J.S. , K. Wang, Y.H. Deng, and G. J. Qi, "PCA-based land use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol.29, no.16, pp. 4823-4838, 2008.

Devereux, B.J. ,G.S. Amable, and C.C. Posada, "An efficient image segmentation algorithm for landscape analysis," *Int. J. Applied Earth Observation and Geoinformation,* vol.6, no. 1, pp. 47-61, 2004.

DigitalGlobe, (2009). *DigitalGlobe* core imagery *products guide.* [Online]. Available: http://www.digitalglobe.com/index.php/6/DigitalGlobe+Products

eCognition, (2010a) . eCognition Developer (8.64.0) User Guide. Trimble Germany GmbH, Trappentreustr. 1, D-80339 München, Germany.

eCognition, (2010b) . eCognition Developer (8.64.0) Reference Book. Trimble Germany GmbH, Trappentreustr. 1, D-80339 München, Germany.

Eikvil, L., L. Aurdal, and H. Koren, "Classification-based vehicle detection in high-resolution satellite images," *ISPRS Journal of Photogramm. Remote Sens.*, vol. 64, no.1, pp. 65-72,2009.

Hinz S. and A. Baumgartner, "Vehicle detection in aerial images using generic features, grouping and context", in *Proc. 23rd DAGM Symp. Pattern Recog.*, Munich, Germany, vol. 2191, pp. 45–52,2001.

Hinz, S., D. Lenhart, and J. Leitloff, "Traffic extraction and characterisation from optical remote sensing data," *Photogrammetric Record* ,vol. 23 (124 SPEC. ISS.), pp.424-440. 2008.

Jin, X., and C. H. Davis, "Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks," *Image Vis. Comput*., vol. 25, no. 9, pp. 1422–1431, Sep. 2007.

Larsen, S. O. , H. Koren, and R. Solberg. "Traffic monitoring using very high resolution satellite imagery," *Photogramm. Eng. Remote Sens.*, vol. 75, no.7, pp. 859–869, 2009.

Larsen S. O. , and A. B. Salberg, "Automatic vehicle counts from Quickbird images," *In Proc. Joint Urban Remote Sens. Event (JURSE),* Munich, Germany, 2011.

Leitloff, J. , S. Hinz, and U. Stilla, "Vehicle detection in very high resolution satellite images of city areas," *IEEE Trans. Geosci. Remote Sens.* vol.48, no.7, pp.2795-2806, 2010.

Liu, W. , F. Yamazaki, and T.T. Vu, "Automated Vehicle Extraction and Speed Determination From QuickBird Satellite Images," *IEEE J. Selected Topics in Applied Earth Observations and Remote Sens.*, vol.4, no.1, pp.75-82, 2010.

Moon, H, R. Chellappa, and A. Rosenfeld, "Performance analysis of a simple vehicle detection algorithm," *Image Vis. Comput.*, vol. 20, no. 1, pp. 1–13, 2002.

Munno, C. J. , H. Turk, J. L. Wayman, J. M. Libert, and T. J. Tsao, "Automatic video image moving target detection for wide area surveillance," *In Proc. Int. Con. Security Technology,* 1993, pp.47-57.

Myint, S. W. , P. Gober, A. Brazel, S. Grossman-Clark, and Q. Weng, "Per-Pixel vs. Object-based classification of urban land cover extraction using high spatial resolution imagery," *Remote Sens. of Environment,*vol 115, pp 1145-1161, 2011.

Nag, S. , and M. Barnes, "A moving target detection filter for an ultra-wideband radar," *In Proc. IEEE Radar Conference*, pp. 147- 153, 2003.

Nie, X. "A semi-automatic framework for highway extraction and vehicle detection based on a geometric deformable model," *ISPRS Journal of Photogramm. Remote Sens.*, vol. 61, pp. 170-186, 2006.

Pesaresi, M. , K.H. Gutjahr, and E. Pagot, "Estimating the velocity and direction of moving targets using a single optical VHR satellite sensor image," *Int. J. of Remote Sens.*, vol. 29, no.4, pp. 1221-1228, 2008.

Radke, R.J. , S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Tran. Image Proc*., vol. 14, no 3. pp. 294-307, 2005.

Ruskone, R. , L. Guigues, S. Airault, O. Jamet, "Vehicle detection on aerial images: a structural approach," *In Proc. 13th International Conference Pattern Recognition*, vol.3, pp. 900-904, 1996.

Salehi, B., Y. Zhang, and V. Dey, "A hierarchical rule based urban land cover classification of WorldView-2 imagery," *In Proc. 32th Canadian Symposium on Remote Sensing*, Sherbrook, Canada, 2011.

Sharma,G., C. J. Merry, P. Goel, M. and McCord, "Vehicle detection in 1-m resolution satellite and airborne imagery," *Int. J. Remote Sens.*, vol. 27, no. 4, pp. 779–797, 2006.

Smiley, B. , "The relative geolocation accuracy of WV02: On-orbit calibration, long term stability," presented at the ASPRS Annual Conference, Milwaukee, Wisconsin, 2011.

Sunar, F. , "An Analysis of Change in a Multi-date Data Set: A Case study in The Ikitelli Area, Istanbul, Turkey". *Int. J. Remote Sens.,*vol. 19, no.2, pp. 225-235, 1998.

Tian, J. , and M. Chen, "Optimization in multi-scale of segmentation of high-resolution satellite images for artificial feature recognition," *Int. J. Remote Sens.,* vol. 28, no.20, pp. 4625-4644,2007.

Toth, C. K., and D. Grejner-Brzezinska, "Extracting dynamic spatial data from airborne imaging sensors to support traffic flow estimation," *ISPRS Journal of Photogram.m Remote Sens*., vol. 61, no. 3/4, pp. 137–148, 2006.

Xiong Z. , and Y. Zhang, "An initial study on vehicle information extraction from single pass of satellite QuickBird imagery," *Photogramm. Eng. Remote Sens*., vol. 74, no. 11, pp. 1401–1411, 2008.

Yamazaki, F., and W. Liu, "Vehicle Extraction And Speed Detection From Digital Aerial Images,"*In Proc. IEEE Symp. Geosci. Remote Sens. (IGARSS 2008),* vol.3, pp. 1334-1337. 2008.

Yao, W., S. Hinz, and U. Stilla, "Extraction and motion estimation of vehicles in single-pass airborne LiDAR data towards urban traffic analysis," *ISPRS J. of Photogramm. Remote Sens.,* vol. 66, no.3,pp. 260-271. 2011.

Zhang, Y. , "The Front Cover: Standard and Colour Enhanced Pansharpening of Quickbird Imagery - Sydney, Australia," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 6. 2004.

Zheng, H., L. Pan, and L. Li, "A morphological neural network approach for vehicle detection from high resolution satellite imagery," *in Proc. Int. Conf. Neural Inform. Process*., vol. 4233, Lecture Notes in Computer Science, I. King, J. Wang, L. Chan, and D. L. Wang, Eds. New York: Springer-Verlag, pp. 99–106,2006.

Zheng, H., and L. Li, "An artificial immune approach for vehicle detection from high resolution space imagery," *Int. J. Comput. Sci. Netw. Secur*., vol. 7, no. 2, pp. 67–72, 2007.

Chapter 6 : SUMMARY AND CONCLUSIONS

This chapter summarizes the research presented in this dissertation. It begins with the summary of each chapter (chapters 2 to 5). The achievements of this research are then presented. Finally, recommendations for future work are provided.

6.1 Summary of Research

In this dissertation the great potential of VHR imagery in urban planning/management is exploited for the following two important applications: land cover classification and moving vehicle extraction. Chapters 2-4 focus on reviewing current techniques and developing new techniques for urban land cover classification, while chapter 5 presents a novel framework for moving vehicle extraction based on the special image characteristics of the newest VHR satellite (WorldView-2).

Chapter 2 presents a comprehensive review of the current literature in the field of urban land cover classification of VHR imagery. Two groups of spatial information that are used for classification are studied, and their usefulness in both pixel-based and object-based classification is discussed. The first group consists of textural, contextual, and morphological (shape, size, extent, etc) measures extracted from the image and the second group includes information from ancillary data such as LiDAR data (height information) and GIS vector data. The review is conducted to assess the usefulness and importance of such measures for urban land cover classification of VHR imagery. An exhaustive list of major publications for each type of information, used in both pixel-

based and object-based approaches, is presented and discussed. In addition, for each type of information, the most frequently used measures (e.g., GLCM texture measures) are presented. The findings of this review provide direction for and justification of research conducted in the subsequent chapters (chapters 3-5)

One of the findings described in Chapter 2 is that, despite the wide availability of GIS vector data of urban areas, such information has not been well utilized in conjunction with VHR imagery for urban land cover classification. Therefore, **Chapter 3** investigated an object-based image analysis framework combining VHR images and GIS height point data. The object-based classification framework is rule-based and is developed using the Cognition Network Language available in the eCognition® software package. In order to test the applicability of the rule-set to different areas and different VHR images, it was developed using a small portion of a QuickBird image and then tested on different and larger QuickBird and IKONOS images. The overall accuracy for QuickBird and IKONOS test images is about 92% and 86%, respectively. Despite the geometric misregistration between VHR images and height point data, these levels of accuracy are very promising for classification of the urban areas with spectrally similar classes such as buildings, roads, and parking lots. This confirms the very high potential of using multisource data in conjunction with object-based image analysis for classification of VHR imagery of urban areas.

Although multisource object-based classification offers promising results for urban land cover classification, its performance depends largely on the availability of ancillary data. In **Chapter 4,** a combined object- and pixel-based image analysis framework is

developed using VHR imagery exclusively. The method starts with segmenting the image using the multiresolution segmentation algorithm, which results in objects with different sizes in different scales (levels of segmentation). Then, for each level of segmentation, several spectral, textural, and morphological features of objects are extracted. To deal with the large number of features, the so-called curse of dimensionality, a novel wavelet-based feature extraction (reduction) method is developed. The wavelet-based feature extraction considerably reduces the number of features, while preserving the dominant information available in the objects' features. Finally, the extracted features together with the original bands of VHR imagery are classified using the Maximum Likelihood algorithm. Results show up to 17%, 10%, and 11% improvement in kappa coefficients compared to the case in which only the original bands of the image are used for WorldView-2, QuickBird, and IKONOS, respectively.

For comparison purposes, the method is tested on WorldView-2, QuickBird, and IKONOS images of the same area acquired on different dates. The wavelet-based feature extraction method is also compared with the conventional PCA-based feature extraction method. In addition, the effects of level of segmentation and type of features on increasing the classification accuracies are quantitatively evaluated.

Chapter 5 presents a novel object-based method for the automatic extraction of moving vehicles (with their speed and direction) using a single WorldView-2 image. The method comprises three major steps: road extraction, moving vehicle extraction, and velocity (speed and direction) estimation. First, roads are extracted using an object-based image analysis framework. Then, the moving vehicles running along the roads are detected and

extracted by combining a PCA-based change detection and object-based image analysis. Finally, the ground positions of moving vehicles were determined using RPC parameters supplied by the satellite, and the velocity and direction of the vehicles are calculated from the difference in the vehicle positions.

6.2 Achievements of the Research

Based on the four main chapters of this dissertation, a summary of the achievements in each chapter is presented as follows:

6.2.1 Review of the Effectiveness of Spatial Information

Until now, no comprehensive study has been performed on the effectiveness of spatial information extracted from the image and from the ancillary data for urban land cover classification of VHR imagery. **Chapter 2** presents a comprehensive review of recent related literature (more than 100 publications including book chapters and peer-reviewed journal papers). Different aspects of the spatial information used for VHR image classification are examined. This chapter (chapter 2) can serve as a comprehensive reference for related studies in the field of VHR image analysis. The major findings of the chapter are:

- The object-based approach generally results in higher classification accuracy than pixel-based approaches when applied to VHR imagery of urban areas. This is because this particular approach takes into consideration the spatial and spectral correlation of a group of neighboring pixels in classification.

- Among all types of spatial information, texture (extracted from the image) and GIS vector data are the most and the least used information in the literature, respectively.

- Textural information is more commonly used in pixel-based classification, while morphological information such as shape and size are is meaningful in object-based classification.

- Contextual information has rarely been used as the only source of spatial information for classification. It has been used in conjunction with textural and/or morphological information. The most commonly used contextual information in the literature is the shadow of high-rise targets such as buildings and trees (e.g., shadow is used for distinguishing buildings from transportation areas).

- In decreasing order, entropy, angular second moment, homogeneity, and contrast of GLCM are the most commonly used textural features in the literature for urban land cover classification of VHR imagery.

- The combination of GIS vector data and VHR images has a great potential for object-based classification of impervious surfaces. The basic processing unit of object-based classification is a group of aggregating pixels which forms an object's polygon and thus can effectively be integrated with vector data. Nevertheless, the integration of vector data and VHR imagery for urban land cover classification is in an early stage and needs further investigation.

6.2.2 Multisource Object-Based Classification

A novel multisource object-based classification method is proposed in **Chapter 3**. The method offers a practical, fast, and easy to use (within eCognition) framework for classifying VHR imagery of small urban areas.

Current literature often ignores two important issues of multisource object-based classification: the transferability of the rule-set to different areas and different VHR images (e.g., transferring a rule-set developed using a QuickBird image to an IKONOS image) and the possible misregistration between different data layers (e.g., between VHR images and vector data). The proposed method mitigates issues mentioned above, while it results in an overall accuracy of up to 92%, which is higher than that reported in the literature. The method also mitigates the effect of possible misregistration between different datasets (which is inevitable in multisource classification) on classification by incorporating objects of higher levels of segmentation into classification. Further, the method shows that it is applicable to different areas of the same image and different VHR images with no change in the rule-set (for the same VHR image) or slight changes in the thresholds (for different VHR images).

Distinguishing spectrally (buildings, roads, and parking lots) and spatially (buildings and parking lots) similar classes in urban areas is very challenging and has not been considered extensively in the current literature. The proposed framework, though it is not applicable to all urban areas, provides guidelines on the types of features (e.g., texture, shape, size, brightness) which are suitable for classifying specific land cover types.

6.2.3 Combined Object- and Pixel-Based Image Analysis

By incorporating the spectral and spatial information of a group of pixels into classification, object-based approaches prove to be effective for urban land cover classification of VHR imagery. However, object-based approaches are not generic and different rules must be developed for different applications. On the other hand, pixel-based approaches are more automatic and generic, but do not consider the information of neighboring pixels. In order to exploit the advantages of object-based and pixel-based approaches, a combined object- and pixel-based image analysis framework is proposed in **Chapter 4**.

The achievements of chapter 4 are as follows:

- A novel and automatic wavelet-based feature extraction method is developed. Current feature extraction techniques are pixel-based and are global in scope. In other words, they consider all the pixels of the image simultaneously and transform the feature space based on a specific criterion (e.g., class separability). Therefore, the computational complexity depends directly on the number of an image's pixels, which is very high in VHR images. As a result, they are not efficient in terms of computational complexity and processing time. As opposed to current techniques, the proposed wavelet-based feature extraction is local in scope and works on the basis of objects not pixels. It takes each object's signature in high dimensional space and transforms it to a lower dimensional space. Therefore, computational complexity depends directly on the number of an image's objects and thus is very efficient in computation and processing time for VHR image analysis (see section 4.4.2.4).

- The chapter demonstrates that additional spectral bands of WorldView-2 (compared to the conventional bands of VHR imagery) do not have a significant effect on improving urban land cover classification.
- It is quantitatively shown that, of spectral, textural, and morphological features of objects, the spectral features have the greatest contribution to distinguishing impervious land cover classes (i.e., buildings, roads, and parking lots). This finding implies that for improving urban land cover classification of VHR imagery, more effort should be directed on using spectral features than using textural or morphological features.
- Incorporating objects of all levels of segmentation into classification is computational-intensive. The chapter investigates the effect of object size (level of segmentation) on classification. It demonstrates that only one level of segmentation is sufficient. It also finds that larger objects (higher levels) are preferred to smaller objects (lower levels) for classifying impervious land cover types. These two last findings have an important impact on improving urban land cover classification of VHR imagery.

6.2.4 Moving Vehicle Extraction

Since 2008, few studies have used the small time lag between the panchromatic and multispectral bands of conventional VHR images, such as QuickBird, IKONOS, and GeoEye-1, to estimate moving vehicle velocity. In these studies, vehicles are either selected manually or semi-automatically using the ancillary road data layer. Thus, the methods are not automatic. In **Chapter 5** the special characteristics of WorldView-2

imagery are exploited, for the first time, to develop a novel and automatic extraction of moving vehicles together with the estimation of the velocity and direction. The method has the following characteristics:

- It uses a single WorldView-2 image exclusively, and road ancillary data, which are often used in the current literature, are not required. Both roads and moving vehicles are automatically or semi-automatically extracted in the proposed method. This achievement has a great impact on automation of moving vehicle extraction using satellite imagery.

- For velocity estimation, the vehicle displacement in panchromatic and multispectral images is required. Current studies often select the vehicle manually in the panchromatic image and then they find the corresponding vehicle in the multispectral image using matching techniques. Then, the vehicle displacement is calculated by subtracting the two positions. Matching, however, is computation-intensive and includes mismatching errors. Our proposed method directly finds the positions of the vehicle in two multispectral sets of a WorldView-2 image (MS1 and MS2) by a change detection technique followed by the object-based extraction of changes. Thus, it avoids computation-intensive matching and its possible errors.

- The time lag between MS1 and MS2 bands of WorldView-2 is almost twice that between QuickBird panchromatic and multispectral bands. As a result, for the same amount of error in a vehicle's positions, the error of speed estimation using QuickBird imagery is twice that of using WorldView-2 imagery.

The method demonstrates a promising potential for automatic and accurate traffic monitoring using a single image of WorldView-2.

6.3 Recommendations for Future Work

Based on the results and contributions discussed in the previous sections, suggestions for future research include the following:

- Although the proposed multisource object-based classification framework provides promising results for urban land cover classification, its performance depends largely on the availability of height point data. This data, however, may not be available for many urban areas. With the availability of stereo VHR imagery in recent years, generating high resolution DEM/DSM for such urban areas is now feasible. Therefore, developing a 3-D object-based classification framework which exclusively uses stereo VHR imagery would be an interesting field of research for the future.

- The combined object- and pixel-based image analysis framework proposed in this research uses the conventional Maximum Likelihood algorithm for classification. Further experiments should be conducted to test the performance of the framework using other classifiers such as Random Forest which has been shown to be effective for classifying VHR imagery.

- For moving vehicle extraction, further research is required to quantitatively evaluate the accuracy of a moving vehicle's positions (i.e., polygons' centroids resulting from segmentation). Since a vehicle's velocity depends on its extracted positions, such research will lead to a quantitative determination of the errors included in the

estimated speeds and directions. Consequently, a more reliable velocity estimation using a single VHR WorldView-2 image could be achieved.

APPENDIX I

Permission from the *International Journal of Geoinformatics:*

Dear Mr. Salehi,

We are thankful for contributing your work in International Journal of Geoinformatics. IJG has the copyright of the paper:

Salehi, B., Y. Zhang, M. Zhong, and V. Dey, 2012. A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery, *International Journal of GeoInformatics*, 8(3):35-51.

It is our pleasure to allow it to be used in your PhD report. Please cite it.

All the best

--
Dr. Nitin Kumar Tripathi
Editor-in-Chief, International Journal of Geoinformatics
Director, UNIGIS Centre, AIT
Associate Dean, School of Engg and Technology
Coordinator, Remote Sensing and GIS
Asian Institute of Technology
P.O. Box:44, Klong Luang, Pathumthani 12120, Thailand
Phone: +66-81751 8384 (Mobile), +66-2-516 2129 (Office),
e-mail: nitingis@gmail.com

APPENDIX II

Permission from *IEEE Journal of Selected Topics in Earth Observation and Remote Sensing:*

Dear Bahram :

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line ©© [year of original publication] IEEE.

2) In the case of illustrations or tabular material, we require that the copyright line © 2012 IEEE appear prominently with each reprinted figure and/or table

3) If you expect to use a substantial portion of the original paper, and if you are not the senior author, please obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © 2012 IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]

2) The published version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.

3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of the University of New Brunswick's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, the Archives of Canada may supply single copies of the dissertation.

In regards to future requests for IEEE copyrighted material, the IEEE has recently implemented a policy that provides an automated means for handling reuse permission of IEEE content through the Copyright Clearance Center RightsLink permission service. You can facilitate this process by taking the steps provided below. If you require answers to specific questions, please contact IEEE Senior Program Manager, Nancy A. Blair-DeLeon. She is responsible for managing this service, and can assist you further. Her contact information is n.blair-deleon@ieee.org

1. Go to http://ieeexplore.ieee.org/Xplore/guesthome.jsp
2. Type the title of the IEEE copyrighted paper into the IEEE Xplore Search Bar and click on the word "Search" .
3. Click on the underlined IEEE copyrighted paper's title that's in orange and blue print.
4. The title will appear again in orange print under the orange Browse tab.
5. Next, click on the orange “Request Permissions” link, which can be found under the IEEE paper title.
6. Continue the process to obtain permission for any material desired from each additional IEEE copyrighted paper.

©©©©©©©©©©©©©©©©©©©©©©©©©©©©©©©©

Jacqueline Hansson, Coordinator
IEEE Intellectual Property Rights Office
445 Hoes Lane
Piscataway, NJ 08855-1331 USA
+1 732 562 3828 (phone)
+1 732 562 1746(fax)
e-mail: j.hansson@ieee.org
IEEE Fostering technological innovation
and excellence for the benefit of humanity.

Curriculum Vitae

Name: BAHRAM SALEHI

Universities Attended:

2008- Present, PhD Candidate, University of New Brunswick, NB, Canada

2005, M.Sc. Remote Sensing, K.N. Toosi University of Technology, Tehran, Iran

2002, B.Sc. Geomatics Engineering, University of Tehran, Tehran, Iran

Publications[1]:

1. **Salehi, B.,** Y. Zhang, M. Zhong (2012). Automatic moving vehicle information extraction from single-pass WorldView-2 imagery. *IEEE Journal of Selected Topic in Earth Observation and Remote Sensing,* Vol 5, No. 1, pp. 135-145.

2. **Salehi, B.,** Y. Zhang, M. Zhong, V. Dey (2012). A review of the effectiveness of spatial information used in urban land cover classification of VHR imagery. *International Journal of GeoInformatics,*Vol. 8, No.3, pp. 35-51.

3. **Salehi, B.,** Y. Zhang, M. Zhong (2012). Object-based land covers classification of urban areas using VHR imagery and Height Points. Journal of *Remote Sensing* , 4(8):2256-2276.

4. **Salehi, B.,** Y. Zhang, M. Zhong (2012). Urban land cover classification using the spectral and spatial features extracted from VHR imagery. Journal of *Photogrammetric Engineering and Remote Sensing* (under review).

5. Dey, V., Y. Zhang, M. Zhong, **B. Salehi** (2011). Image segmentation techniques for urban land cover segmentation of VHR imagery: Recent Developments and Future prospects. *ISPRS Journal of Photogrammetry and Remote Sensing* (under review).

6. Dey, V., Y. Zhang, M. Zhong, **B. Salehi** (2011). A Supervised Methodology for Optimal Parameter Estimation of Multiresolution Segmentation within eCognition. *Canadian Journal of Remote Sensing* (under review).

---

[1] The following publications are those achieved during my PhD study at University of New Brunswick. Publications achieved prior to September 2008 are not included in this list.

7. **Salehi, B.,** Y. Zhang, M. Zhong (2012). Object-based moving vehicle extraction from WorldView2 imagery. *In Proc. of ASPRS 2012 Annual Conferenc*e, Sacramento, California.

8. **Salehi, B.,** Y. Zhang, M. Zhong (2012). The effect of four new multispectral bands of WorldView2 on improving urban land cover classification. *In Proc. of ASPRS 2012 Annual Conferenc*e, Sacramento, California.

9. **Salehi, B.,** Y. Zhang, R. Mishra (2011).Automatic Moving Vehicle Extraction and Speed Determination from Single Pass WorldView-2 Multispectral Bands. In the Archive of The DigitalGlobe 8-band Research Challenge. Online available: http://dgl.us.neolane.net/res/dgl/survey/_8bandchallenge_researchPapers.jsp?deliveryId=&id=.

10. **Salehi, B.,** Y. Zhang, M. Zhong (2011). Object based Land cover Classification of Urban Areas Using VHR Imagery and Photogrammetrically Derived DSM. *In Proc. of ASPRS 2011 Annual conferenc*e, Milwaukee, Wisconsin.

11. **Salehi, B.,** Y. Zhang, M. Zhong (2011). Combination of pixel based and object based image analysis for classification of VHR imagery over urban areas. *In Proc. of ASPRS 2011 Annual conferenc*e, Milwaukee, Wisconsin.

12. **Salehi, B.,** Y. Zhang,M. Zhong, V. Dey (2011). A hierarchical rule based urban land cover classification of WorldView-2 imagery. *In Proc. of 32th Canadian Symposium on Remote Sensing*, Sherbrook, Canada.

13. 15. Dey, V., Y. Zhang, M. Zhong, **B. Salehi** (2011). Building detection using multi-level segmentation with a fuzzy parameter based region merging criteria. *In Proc. of 32th Canadian Symposium on Remote Sensing*, Sherbrook, Canada.

14. Beykaei, S.A, M. Zhong,Y. Zhang, **B. Salehi**, M. Ircha, Y. Gweon, S. Gao (2010). Evaluating the accuracy of extracting urban land cover/use from remotely sensed imagery and its potential application in urban planning. *In Proc of Annual Conference - Canadian Society for Civil Engineering* 2, pp. 1539-1548.

15. Zhong, M., Y. Zhang, S.A. Beykaei , **B. Salehi**, M. Ircha, Y. Gweon, S. Gao, (2010). Extracting land cover/use from remotely sensed imagery: Potentials for urban planning. *In Proc. of the Conference on Traffic and Transportation Studies, ICTTS* 383, pp. 1-12.

Conference Presentations[1]:

[1] The following presentations were made during my PhD study at University of New Brunswick. Presentations made prior to September 2008 are not included in this list.

Salehi, B., Y. Zhang, M. Zhong, 2012. Wavelet-based reduction of objects' features for classification of VHR imagery of urban areas. *Global Geospatial Conference*, Quebec City, Quebec, CANADA.

Salehi, B., Y. Zhang, M. Zhong, 2012. Object-based moving vehicle extraction from WorldView2 imagery. *ASPRS 2012 Annual Conferenc*e, Sacramento, California. USA.

Salehi, B., Y. Zhang, M. Zhong, 2012. The effect of four new multispectral bands of WorldView2 on improving urban land cover classification. *ASPRS 2012 Annual Conferenc*e, Sacramento, California, USA.

Salehi, B., Y. Zhang,M. Zhong, V. Dey, 2011 A hierarchical rule based urban land cover classification of WorldView-2 imagery. *32th Canadian Symposium on Remote Sensing*, Sherbrook, Quebec, CANADA.

Salehi, B., Y. Zhang, M. Zhong, 2011. Object based Land cover Classification of Urban Areas Using VHR Imagery and Photogrammetrically Derived DSM. *ASPRS 2011 Annual conferenc*e, Milwaukee, Wisconsin, USA.

Salehi, B., Y. Zhang, M. Zhong, 2011. Combination of pixel based and object based image analysis for classification of VHR imagery over urban areas. *ASPRS 2011 Annual conferenc*e, Milwaukee, Wisconsin, USA.

Chairing Technical Sessions

- **Technical Challenges** at *Global Geospatial Conference 2012*. The joint *GSDI World Conference (GSDI 13),* the *14th GEOIDE Annual Scientific Conference***,** the *7th 3DGeoInfo Conference***,** and the *Canadian Geomatics Conference.* Quebec city, Canada, May 14-17 2012.
- **Object-based Image Analysis** at *American Society of Photogrammetry and Remote Sensing (ASPRS) Annual Conference*. Sacramento, California, March 19-23, 2012.
- **Land Cover/Land Use** at *American Society of Photogrammetry and Remote Sensing (ASPRS) Annual Conference*. Sacramento, California, March 19-23, 2012.