AUTOMATED EXTRACTION OF BUILDINGS FROM IKONOS IMAGERY BY INTEGRATING SPECTRAL AND SPATIAL INFORMATION

TK
8315
W36
2006

By

Xin Min Wang

B.Eng, Tongji University, China, 1992

A thesis presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Applied Science

in the Program of Civil Engineering

Toronto, Ontario, Canada, 2006

© Xin Min Wang 2006

UMI Number: EC53548

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform EC53548
Copyright 2009 by ProQuest LLC
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Xin Min Wang

~

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Xin Min Wang

~

Department of Civil Engineering

Ryerson University

BORROWER’S PAGE

Ryerson University requires the signatures of all persons using or photocopying this thesis.

Please sign below, and give address and date.

Name	Address	Date	Signature

Automated Extraction of Buildings from IKONOS Imagery by Integrating Spectral and Spatial Information

by

Xin Min Wang

A Thesis Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements for the Degree of Master of Applied Science

in the Program of Civil Engineering at Ryerson University

Toronto, Canada, 2006

ABSTRACT

Very High Resolution (VHR) imagery is becoming indispensable in urban mapping and land management. But owing to its characteristics, traditional spectral-based building extraction methods are not appropriate. This thesis presents an automated spectral-spatial integration extraction approach to extract two-dimensional roof outlines of urban buildings from IKONOS imagery. It substantially improves accuracy, compared with the well-known Extraction and Classification for Homogenous Objects (ECHO) and Iterative Self-Organizing Data Analysis Technique (ISODATA). With spectral-based Multi-peak Supervised Segmentation (MSS), the

approach generates first stage classification. In the second stage, the original colour image and the first stage result are fed into a spatial structure classifier. With analysis of overlapped neighborhood's degree of membership to classes, homogeneity and cross-neighborhood structure, the first stage classification is significantly improved. Finally, buildings are extracted with post-processing. Compared with ISODATA, Signature Editor and ECHO, unique contributions are three new algorithms: MSS, Overlapped Neighborhood Analysis, and Cross-neighborhood Structure Detection.

ACKNOWLEDGEMENTS

I express my sincere gratitude and appreciation to many people who have made this thesis possible.

First and foremost, I would like to thank my supervisor, Professor Dr. Jonathan Li, for his great support and help in my study to pursue this goal lasting over thirty years. Nothing is more important than the values and principles he taught me. Thanks also go to Dr. Kunquan Lan for serving as my co-supervisor and for his critical comments and encouragement.

I appreciate all those who have helped me in my work at Ryerson University, Professors Dr. Michael A. Chapman, Dr. Songnian Li, and Dr. Resa Kanoish, my fellow graduate students Mr. Yu Li, Mr. Xiangyu Si and Mr. Robin Wan. I have learned a great deal from them during my time here. Special thanks also go to Ms Ellen Dale for her proofreading and suggestions improving this thesis.

I also would thank my wife and my family members, who have made my life colourful and wonderful. They have helped and encouraged me a lot throughout my study, and finally helped my dreams come true. I really appreciate what they have done for me for so many years, and look forward to sharing my success and happiness with them.

TO MY BELOVED

Lei Zhang

TABLE OF CONTENTS

DECLARATION.....	II
BORROWER’S PAGE.....	III
ABSTRACT	IV
ACKNOWLEDGEMENTS.....	VI
DEDICATION.....	VII
TABLE OF CONTENTS.....	VIII
LIST OF TABLES	XII
LIST OF FIGURES	XIII
ABBREVIATIONS	XVIII
CHAPTER 1: INTRODUCTION.....	1
1.1 NEEDS FOR HIGH-RESOLUTION SATELLITE IMAGERY FOR URBAN MAPPING.....	1
1.2 2D BUILDING ROOF EXTRACTION.....	2
1.3 IMAGE SEGMENTATION AND CLASSIFICATION: EXISTING TECHNIQUES AND CHALLENGES	4
1.4 MOTIVATIONS, OBJECTIVE AND SCOPE OF STUDY	9
1.5 THESIS ORGANIZATION	13

CHAPTER 2: LITERATURE REVIEW	15
2.1 IMAGE SEGMENTATION AND CLASSIFICATION METHODS	16
2.2 SPECTRAL-BASED SEGMENTATION ALGORITHMS	17
<i>2.2.1 Minimum Distance to Means.....</i>	<i>18</i>
<i>2.2.2 K-Means.....</i>	<i>19</i>
<i>2.2.3 Fuzzy C-Means.....</i>	<i>21</i>
<i>2.2.4 ISODATA.....</i>	<i>22</i>
<i>2.2.5 Drawbacks of Spectral-based Algorithms.....</i>	<i>25</i>
2.3 TEXTURE-BASED SEGMENTATION ALGORITHMS.....	26
<i>2.3.1 Markov Random Field and Model-based Segmentation</i>	<i>27</i>
<i>2.3.2 Discussion</i>	<i>29</i>
2.4 OBJECT-ORIENTED SEGMENTATION ALGORITHMS.....	30
2.5 SPECTRAL-SPATIAL SEGMENTATION ALGORITHMS	33
<i>2.5.1 Spectral-spatial Sequential Integration.....</i>	<i>34</i>
<i>2.5.2 Weighted Additive Combination</i>	<i>42</i>
<i>2.5.3 Integrative Colour-texture Vectors</i>	<i>45</i>
2.6 SUMMARY OF EXISTING METHODS	46
CHAPTER 3: PROPOSED METHODOLOGY	49
3.1 INTRODUCTION.....	49
<i>3.1.1 RGB Colour Space</i>	<i>49</i>

3.1.2 3D Colour Feature Distribution (<i>Histogram</i>)	51
3.1.3 Morphological Operation.....	54
3.2 FLOWCHART OF PROPOSED METHODOLOGY	58
3.3 DESCRIPTION OF ALGORITHMS	61
3.3.1 Supervised Segmentation Logic	62
3.3.2 Multi-peak Supervised Segmentation	63
3.3.3 Integration of Spectral and Spatial Information	68
3.3.4 Morphological Filtering.....	78
3.3.5 Small Object Filtering.....	79
3.3.6 Building Boundary Extraction and Overlapping.....	79
3.4 CHAPTER SUMMARY	80
CHAPTER 4: IMPLEMENTATION.....	82
4.1 DATASETS AND SOFTWARE ENVIRONMENT	82
4.2 EXPERIMENTS.....	84
4.2.1 Spectral Segmentation Comparison between MSS and Unsupervised ISODATA	84
4.2.2 Spectral segmentation comparison between MSS and Signature Editor (SE)	88
4.2.3 Comparison between Spectral-only and Spectral-spatial Classification.....	92
4.2.4 Comparison between proposed spectral-spatial approach and ECHO	96
4.2.5 Comparison between Different Neighborhood Sizes	99
4.2.6 Comparison between Different Thresholds.....	104

4.3 CHAPTER SUMMARY	109
CHAPTER 5: RESULTS AND QUALITY EVALUATION	110
5.1 GENERATION OF GROUND TRUTH IMAGES	110
5.2 EVALUATION METHODOLOGY	111
5.3 EVALUATION OF RESULTS AND DISCUSSION	115
5.3.1 <i>Confusion Matrix and Kappa Statistics</i>	115
5.3.2 <i>Building Extraction Rate</i>	126
5.3.3 <i>Factors Affecting the Accuracy of Results</i>	131
CHAPTER 6: CONCLUSIONS AND RECOMMODATIONS	132
6.1 CONTRIBUTIONS	132
6.2 RECOMMENDATIONS FOR FUTURE RESEARCH	134
REFERENCES	136
APPENDIX A	149
APPENDIX B	164
APPENDIX C	168

LIST OF TABLES

Table 2- 1. The summary of existing methodologies of multispectral imagery segmentation.....	47
Table 3- 1. Rules for grayscale dilation and erosion.....	55
Table 5- 1. A confusion matrix example (unit: pixel).....	112
Table 5- 2. Confusion matrix of ISODATA, SE, MSS results	117
Table 5- 3. Confusion matrix of ISODATA, SE, MSS results	119
Table 5- 4. The confusion matrix of spectral-only and spectral-spatial segmentation	121
Table 5- 5. The confusion matrix of ECHO, ECHO+ ONA, ECHO+ CSD, proposed approach classification results	123
Table 5- 6. Summary of Overall Accuracy and the κ evaluation	125
Table 5- 7. The extraction rate of ISODATA, SE, and MSS.....	127
Table 5- 8. The extraction rate of ECHO and proposed approach.....	129
Table 5- 9. The extraction rate of ECHO and proposed approach.....	130
Table 5- 10. The summary of extraction rate comparison	130

LIST OF FIGURES

Figure 2- 1. Multispectral segmentation and classification groups: spectral-based, texture-based, object-oriented, and spectral-spatial approaches.	16
Figure 2- 2. Schematic diagram illustrates the classification flow in eCognition object-oriented approach (Wong et al., 2005).	31
Figure 2- 3. The concept of scale, in which mainly three different levels of image objects have been created representing different scales (Wong et al., 2005).....	32
Figure 2- 4. Flowchart of the granite defect detection	36
Figure 2- 5. (a) Data are poorly represented by clusters A and B; (b) super clusters A and B are formed by many sub-clusters. (Song et al., 1996).....	37
Figure 2- 6. Segmentation of a HYDICE hyperspectral image (Jimenez et al., 2005).....	40
Figure 2- 7. (a) Result of C-means algorithm applied at first stage, (b) Result of UnECHO algorithm applied at the second stage. (Jimenez et al., 2005)	41
Figure 2- 8. Region-based segmentation; split-merge plus pixel-wise refinement: (a) the workflow, (b) region similarity measure is critical. (Hu et al., 2005).....	43
Figure 3- 1. The RGB colour space and cube.	50
Figure 3- 2. Discretization of RGB colour space and codes of bins.....	53
Figure 3- 3. A 3D colour feature distribution (histogram) with 512 bins (the green dots represent the major distribution peaks).	54

Figure 3- 4. Morphological Processing of a Binary Image.....	56
Figure 3- 5. Original building image (left); and building roof extracted in binary image (right).	57
Figure 3- 6. The ‘opening’ morphological operation: erosion (left) + dilation (right).	57
Figure 3- 7. The ‘closing’ morphological operation: dilation (left) + erosion (right).....	58
Figure 3- 8. Flowchart of spectral-spatial integration.	59
Figure 3- 9. 100×79 ‘BUSH’ image (top right) has 5 prominent peaks covering $\geq 70\%$ pixels..	65
Figure 3- 10. 37×36 ‘ROOF’ image (top right) has 9 prominent peaks, covering $\geq 60\%$ pixels.	65
Figure 3- 11. 56×24 pixels ‘BARELAND’ image (top right) has only 4 prominent peaks, but covers $\geq 90\%$ pixels.	66
Figure 3- 12. 3 × 6 pixel grid with non-overlapped 3 × 3 neighborhoods (left), and with overlapped 3 × 3 neighborhoods (right).	71
Figure 3- 13. (a) Various scenarios of neighborhood’s locations (type A and type B) around a building (shaded area). (b) The final classification result of the building: many misclassified concave patches and convex patches.....	74
Figure 3- 14. A targeted building pixel stretches out four STINGs to detect its cross-neighborhood homogeneity.	76
Figure 3- 15. A targeted non-building pixel stretches out four main STINGs and four secondary STINGs to detect its cross-neighborhood homogeneity.....	77
Figure 3- 16. Original image (a), buildings before morphological filtering (b), and buildings after morphological filtering (c)	78

Figure 3- 17. Before small object filtering (notice the spots besides arrows) (left); after small object filtering, small impurities disappeared (right).	79
Figure 3- 18. Extracted building roof outlines (left), and building roof outlines superposed on original image to visually check the extraction accuracy (right).	80
Figure 4- 1. Four test images used in the study (a) townhouses, (b) community centre, (c) industrial and (d) apartment buildings.....	83
Figure 4- 2. Spectral segmentation comparison. Original images (a), Results of ISODATA (b), and Results of MSS (c).....	86
Figure 4- 3. Further comparison between ISODATA and MSS. Original images (a); Spectral-spatial integration classification based on ISODATA spectral result (b); Spectral-spatial integration classification based on MSS spectral result (c).....	88
Figure 4- 4. Comparison between SE and MSS. Original images (a); SE results (b); MSS results (c).	90
Figure 4- 5. Further comparison between SE and MSS. Original images (a); Spectral-spatial integration classification based on SE spectral result (b); Spectral-spatial integration classification based on MSS spectral result (c).	92
Figure 4- 6. Comparison between spectral-only and spectral-spatial integration. Original images (a); Spectral-only extractions (b); Spectral-spatial extractions (c); Spectral-only extractions and ground truth (d); Spectral-spatial extractions and ground truth (e).	95
Figure 4- 7. Comparison between ECHO and proposed integration approach. Original images (a);	

ECHO building extractions (b); Proposed approach extractions (c); ECHO extractions and ground truth (d); Proposed approach extractions & ground truth (e).....	99
Figure 4- 8. Different accuracy with different neighborhood sizes in “Community centre”. Final extraction results in Column (a), and final results superposed on original images in column (b).	101
Figure 4- 9. Different execution time with different neighborhood sizes.....	101
Figure 4- 10. Different accuracy with different neighborhood sizes in “Storage buildings”. Final extraction results in Column (a), Final results superposed with original images in column (b).	103
Figure 4- 11. Different execution time with different neighborhood sizes.....	103
Figure 4- 12. Threshold coefficient comparison in “Community center”.	106
Figure 4- 13. Threshold coefficient comparison in “Storage buildings”.	108
Figure 5- 1. Ground truth examples. (a) & (b) Townhouses and its ground truth, (c) & (d) Industrial buildings and its ground truth	111
Figure 5- 2. “Townhouses” spectral-based classification comparison among MSS, ISODATA and SE. Results with ground truth (a), and results superposed on original image (b).	116
Figure 5- 3. “Industrial buildings” spectral-based classification comparison among MSS, ISODATA and SE. Results with ground truth (a), and results superposed on original image (b).	118
Figure 5- 4. Spectral-only and spectral-spatial results. (a). Results with ground truth (in red) (b);	

and results (in blue) superposed on original image (c).....	120
Figure 5- 5. ECHO, ONA, CSD and the proposed approach comparison. Results with ground truth (in red) (a), and results (in blue) superposed on original image (b).	123
Figure 5- 6. Spectral comparison among MSS, ISODATA and SE. Results compared with ground truth (a); results superposed on original image (b).	127
Figure 5- 7. Comparison between ECHO and proposed approach in “Townhouses”. Results compared with ground truth (a); Results (in blue) superposed on original image (b).	128
Figure 5- 8. Comparison between ECHO and proposed approach in “Industrial buildings”. Results compared with ground truth (a); Results superposed on original image (b).	129

ABBREVIATIONS

2D	two dimensional
3D	three dimensional
CSD	Cross-neighborhood Structure Detection
DCS	digital camera system
ECHO	Extraction and Classification for Homogenous Objects
FCM	Fuzzy C-Means
GIS	Geographic Information System
GPS	Global Positioning System
ISODATA	Iterative Self-Organizing Data Analysis Technique
MLC	Maximum Likelihood Classifier
MRF	Markov Random Field
MSS	Multi-peak Supervised Segmentation
ONA	Overlapped Neighborhood Analysis
RGB	red, green, blue
SAR	Synthetic Aperture Radar imagery
TS-MRF	Tree Structure Markov random field
UnECHO	Unsupervised Extraction and Classification for Homogenous Objects
VHR	very high resolution

Chapter 1

INTRODUCTION

1.1 Needs for High-resolution Satellite Imagery for Urban Mapping

Spatial information is of great importance for urban and transportation planning, land development and macro-economy in urban areas. A number of vital applications increasingly rely on more detailed and accurate spatial information of buildings and roads in complicated urban areas, for instance, urban disaster response and management, urban heat island effect, city sprawl change detection and geographic information systems (GIS) database update.

However, the acquisition of such detailed and accurate spatial information from conventional aerial photographs, used by city planners and managers, is a very time consuming and expensive process. Nowadays, very high resolution (VHR) satellite imagery has received considerable attention due to its unique advantages.

Compared with aerial photographs, VHR satellite imagery provides current information on a periodic basis at low cost and allows for extraction of more detailed and accurate information which is difficult to be extracted from medium-to-coarse resolution satellite imagery (e.g., 30 m Landsat-5 Thematic Mapper imagery). VHR satellite imagery, with an instantaneous field of

view (IFOV) of 1 m^2 or less, presents an opportunity to develop entirely new approaches to colour image analysis (Warner and Steinmaus, 2005). Moreover, VHR satellite images (e.g., QuickBird, IKONOS, SPOT 5, and Orbview-3) increase the amount of information available on land cover, land use, hydrologic information, and disaster management data. Therefore, they are becoming a prominent data source for urban geospatial applications (Hu et al., 2005).

1.2 2D Building Roof Extraction

As is well-known, the sensory data rarely contains information in an explicit form since they are raw data. It is impossible to directly use them without information extraction. Accurate and efficient information extraction is the first and key step in spatial data applications. However, this is one of the difficulties in the digital image processing. After all, it is the processing and analysis of data that constitute a real bottleneck, not the gathering of data (Schenk, 2003).

Two-dimensional (2D) building extraction involves understanding an image and extracting man-made structures from raw or preprocessed sensor data, and therefore get a certain likelihood of building's physical properties in digital form, such as the shape, orientation or boundary. 2D building extraction lays a solid foundation for the later three-dimensional (3D) building reconstruction, which is the goal of digital photogrammetry and computer vision (Grun et al., 1995; Grun et al., 1997; Haala and Brenner, 1999).

The core technology in 2D building extraction involves the image segmentation and classification process.

Segmentation is the automated process of partitioning an image into several clusters that are homogeneous with respect to some characteristics like colour, texture, reflection signal, or context, etc. (Chen et al., 2002). Classification is the process of assigning segmented individual pixels or homogeneous clusters to specific and more meaningful information classes by human supervision. The main objective of image classification is to automatically categorize all pixels in an image into land-cover classes or themes, or to distinguish maximum homogeneous regions to different objects (Asano et al., 1996).

Satellite images are generally available in a discrete digital format and represent the spatial distribution of energy emanating from the earth's surface in a sensible range of wavelengths. Consequently, VHR satellite images are usually originated in multivariate forms such as multispectral and multitemporal images. As the satellite's spatial resolution increases, the multivariate (internal variability) increases too (Carleer et al., 2005). As a direct result, processing more multivariate images is more complicated due to the increased dimensions and the need for extracting and exchanging information from and among all components. This reduces segmentation and classification accuracy. Additionally, VHR satellite imagery has not

yet reached the same high spectral resolution as aerial images, although they have the advantages of high spatial resolution, a short revisit period and lower cost. With relatively low spectral resolution, the VHR satellite imagery has limited spectral capabilities (Carleer et al., 2005). These aspects make the task of segmentation and classification from VHR satellite imagery difficult and complex.

Moreover, compared with rural areas, urban areas are more heterogeneous than natural landscapes. In urban areas buildings of different sizes, shapes, colours and textures are constituted of more diverse man-made materials mixed with natural materials, like concrete, steel, asphalt, brick or plastics, and crowded into small spaces. Therefore, buildings often blend into their environment, and this makes the spectral extraction method not very effective. In addition, many high-rise shadows are also disturbing to the building roof detection process.

Therefore, more sophisticated segmentation and classification techniques are needed for extracting and interpreting the urban building information captured by VHR satellite imagery.

1.3 Image Segmentation and Classification: Existing Techniques and Challenges

The segmentation and classification is a fundamental technique in the thematic information extraction from remote sensing imagery, and has been a long-standing bottleneck in satellite

image applications, as well as a very active research topic in digital image processing. Although there are some differences between the definitions of segmentation and classification as aforementioned, in general, image segmentation and classification technique can be subdivided into spectral-based, texture-based, object-oriented, and spectral-spatial integration approaches.

Spectral-based approaches

Earlier segmentation methods focused on relatively simple multispectral properties, for example, ellipse plots (Toll et al., 1977), maximum likelihood classification (Swain et al., 1978), and fuzzy sets (Zadeh, 1965), all made good use of the pixel's RGB(red, green, blue) values. Various spectral-based (pixel-spectral) algorithms have been developed, such as the algorithms based on parametric and nonparametric statistics, the use of supervised or unsupervised classification logic, the use of hard or soft (fuzzy) set classification logic to create hard or fuzzy thematic output products.

The results obtained from these methods are not satisfactory. The typical defect of the “salt-and-pepper” effect drew most criticism, particularly, in the case of mapping anthropogenic structures and other heterogeneous features in complex urban scenes (Kiema, 2002). Pixel-based segmentation and classification is often characterized by limited accuracy and low reliability (Haala and Brenner, 1999). Insufficient spatial resolution of the satellite sensors and insufficient

use of the spatial or structural information in the classification procedure are the two main reasons proposed to explain this failure (Cushnie, 1987).

Texture-based approaches

Haralick and Bryant (1976) gave image texture a definition as “the pattern of spatial distributions of gray-tone”. From the visual point of view, texture is described as a pattern with some kind of regularity to characterize the spatial interaction of spectrum. In particular, texture reflects the spatial structure of pixel values, and it is, therefore, indispensable in segmenting an area into sensible geographical units (Lucieer et al., 2005).

The approaches of texture mathematical modeling (gray-tone) are grouped into structural, statistical and signal theory methods (Reed and DuBuf, 1993). The texture segmentation strategy for 2D building extraction falls into two classes: (1) edge-based (i.e., linear feature extraction based on discontinuities in an image), and (2) region-based (i.e., area feature extraction based on the similarities in an image).

The results obtained from above texture approaches are barely satisfactory, particularly for VHR satellite imagery due to its relatively lower spectral resolution. Moreover, when considering the texture features, almost all the images are dealt with in gray tone, i.e., the spatial texture

information was extracted independently in single panchromatic band or gray level imagery, with little interactive relation to multispectral information.

Object-oriented approaches

In contrast to pixel-based spectral methods, another major category is the object-oriented (region-growing) approach. It creates individual areas with shape and spectral homogeneity, which one may recognize as segments or patches. The process can be clearly split into two steps, segmentation and classification. Objects were assigned class rules using spectral signatures, shape and contextual relationships. The rules were then used as a basis for fuzzy classification (Whiteside and Ahmad, 2005).

The object-oriented approach is generally better than feature space clustering or pixel based approaches, and utilizes region homogeneity statistically and so tends to be less sensitive to noise. However, it is not suitable for building extraction, because it is difficult to assign correct weights for multi-sized building objects during the segmentation process, and also easy to create problems in defining the segmentation scale level for a particular object, i.e., missing buildings after classification process is a problem (Mayunga, 2006). Furthermore, the uncertainty of segmentation output in the building roof outlines (i.e., transition zones) is still unsatisfactory (Whiteside and Ahmad, 2005; Carleer et al., 2005).

Spectral-spatial integration approaches

Mainly after the new millennium, people came to realize the defects of only using texture information or only using spectral information, and started to combine the spatial texture or intensity information with spectral information. This resulted in a significant improvement in image segmentation and classification, and this methodology is generally called a spectral-spatial integration approach. Currently this is the most popular and effective approach for building extraction with the growing use of VHR satellite imagery.

There are three typical strategies for the spectral-spatial integration approach; namely (1) spectral-spatial sequential integration (Jimenez et al., 2005; Mirmehdi and Petrou, 2000), (2) weighted additive combination (Caelli and Reye, 1996; Hu et al., 2005; Eom, 1999), and (3) integrative colour-texture vectors (Lucieer et al., 2005; Paschos and Petrou, 2003; Ilgner et al., 2003; Palm, 2004).

Challenges in segmentation and classification

Just as Davies (1990) stated in machine vision, “It must not be overlooked that segmentation is actually one of the central and most difficult practical problems of machine vision”, similarly, in 2D building extraction, most of the difficulties encountered also lie in accurate image

segmentation or classification. The general problem is separating buildings or building edges from their background. For example, if the surface reflectance of the object is not uniform from the background and the image scene is not evenly illuminated, then the resulting building will not be a homogeneous object. Or, if the object's material has similar colour and texture to its background, it is difficult to separate it away from the environment, even though the spectral-spatial integration approach is used.

1.4 Motivations, Objective and Scope of Study

Based on the above segmentation and classification problems encountered in current 2D building extraction, the following aspects represent the original motivation for commencing this study:

Firstly, the role of VHR satellite imagery in supporting the urban planning and land management is increasingly and extensively being recognized by governments, communities and non-government organizations (NGOs). However, due to more multivariate and relatively low spectral resolution, extraction accuracy from VHR satellite imagery is still not improved enough when compared with aerial imagery, which remains much more costly. Consequently, more effective and practical segmentation and classification approaches for VHR satellite imagery are needed.

Secondly, a new supervised segmentation must be used as a prominent key to improve current segmentation accuracy. In building extraction, conventionally, unsupervised segmentation was usually used. But when the surface reflectance of the object is not uniform from the background and the image scene is not evenly illuminated, the resulting unsupervised segmented image will be a mess; and if a building has multi-material or has similar colour and texture to its background, separating it from the background is very difficult, even though the colour-texture combined approach is used. However for human eyes, there are no such problems; it is easy for them to tell multi-colour, multi-material or uneven reflected buildings from their background. Why? Experience does the job! Nevertheless, supervised classification can give the descriptor good experience. Supervised classification is the process of using priori samples of known identity (feature signatures) to classify unknown objects with the human experience, therefore it is a promising approach to solve this problem, including the multivariate weakness in VHR satellite imagery. But conventional supervised classification has an annoying obstacle in priori sample (signature) selection: their final results are very sensitive to the priori samples, which results in strong bias. Thus, an innovative supervised classification is needed.

Thirdly, among so many current approaches, the spectral-spatial sequential integration approach is proven straightforward and effective in building extraction (Jimenez et al., 2005; Mirmehdi and Petrou, 2000; Gong and Howarth, 1990; Marceau et al., 1990). From the delineation of segmentation and classification technology in the previous sections, it is obvious to see that both

spectral-based and texture-based segmentation and classification have their fatal defects, as a result, the integration of spatial and spectral information is a right direction to solve problems. Unfortunately, these approaches have been tested on medium resolution satellite imagery (e.g., 10-20 m SPOT-4 imagery), but they have not been applied on VHR satellite imagery (Puissant et al., 2005). This is another main motivation of this study to exploit existing spectral-spatial integration approaches to improve the building roof extraction from VHR satellite imagery.

The objective of this study is to develop a spectral-spatial integration approach to extract 2D urban buildings from 1 m IKONOS imagery. More specifically, the proposed approach is a spectral-spatial sequential integration. With the Multi-peak Supervised Segmentation (MSS) based on 3D colour feature distribution, this automated extraction approach will generate spectral-based classification. The original colour image and the result of spectral-based classification are then fed into a spatial structure classifier. With analysis of the neighborhood's degree of membership to classes, homogeneity and cross-neighborhood structure, the previous classification is revised and improved. Finally, after post-processing, buildings and their boundaries are extracted. The proposed MSS method is an innovative and effective algorithm owing to its easy signature sample selection and high degree of accuracy. The spatial structure classifier is an innovative well development of the well known Extraction and Classification for Homogenous Objects (ECHO) classifier (Jimenez et al., 2005; Landgrebe, 1980; Richrds, 1993), in which two new algorithms are exploited: one is the overlapped neighborhood structure instead

of non-overlapped neighborhood structure, another is the innovative cross-neighborhood structure detection. To the end, experiment results will demonstrate that this sequential approach is much more effective than the ones combining the Iterative Self-Organizing Data Analysis Techniques (ISODATA) segmentation with same spatial classifier or the same MSS approach but UnECHO classifier.

Two factors limit the scope of this study. First, as stated in Section 1.3, while dealing with the texture spatial information, there are two segmentation directions in automated 2D building extraction: edge-based (linear feature extraction) or region-based (area feature extraction). Typical approaches are Optimal Edge Detector based on Gaussian filter (Canny, 1986; Cocquerez and Philipp, 1995), Laplacian of Gaussian (LOG) operator (Marr and Hildreth, 1980), Watershed Segmentation (Vincent and Soille, 1991; Debeir, 2001), Binary Space Partitioning Tree and so on. In common, edge-based approaches do not lead directly to a segmentation of the image because obtained building contours are seldom closed. Moreover, they only work in gray tone, i.e., the spatial texture information was extracted independently in single panchromatic band or gray level imagery, with little interactive relation to multispectral information, therefore, they are very sensitive to noise or texture in the images (Carleer et al., 2005), which is their fatal limitation and results in low accuracy so far. To this end, the development presented in this thesis is concentrated on the more popular region-based approaches in current trends to enhance the accuracy for building roof extraction. Region-based approaches may be the only one able to

bring a significant contribution to production in the near future.

Second, in order to have a good comparison with current related studies and popular satellite images, this study focuses on orthophotography in natural colour from 1m resolution IKONOS imagery, applying to multispectral data extraction.

1.5 Thesis Organization

This thesis consists of following six chapters:

Chapter 1 provides the background information of 2D building extraction from high-resolution satellite imagery, and briefly reviews the status of segmentation and classification technology. Then, the motivation, objective, scope of this study and the structure of the thesis are described.

Chapter 2 reviews the segmentation and classification methodologies, organizes and analyzes recent publications pertaining to multispectral satellite imagery. Emphasis is given to comparisons among them and their applicability to the 2D building extraction in urban area from VHR satellite imagery.

Chapter 3 presents theoretical background and then the framework and flowchart of proposed

methodology. Following that is the detailed description of proposed methodology: the Multi-peak Supervised Segmentation and improved sequential spectral-spatial integration classification.

Chapter 4 describes the implementation undertaken on the new classification approach. The comparison and discussion between proposed algorithms and other algorithms, including ISODATA and UnECHO, are reported in the experiment results.

Chapter 5 presents qualitative and quantitative evaluation methodologies for the experiment results, and then focuses on experiment result evaluation and discussion.

Chapter 6 summarizes the contributions of the research work, and recommends research topics of future work.

Chapter 2

LITERATURE REVIEW

The goal of this chapter is to review recent publications pertaining to segmentation and classification methods for 2D building extraction from multispectral satellite imagery. The delineation of the task of 2D building extraction will be presented first. Then, as an overview of segmentation and classification algorithms, emphasis is given to the comparison among them and their applicability to the 2D building extraction in urban area from VHR satellite imagery. The problems around the issues pertaining to these algorithms are also discussed.

The task of 2D building extraction is to extract the building shape, orientation, boundary, area or other properties from raw or preprocessed sensor data, to complete mapping or engineering projects. It is fundamental for the later automated 3D building reconstruction, which is imperative in current urban disaster management, environment monitoring, and GIS applications.

The core process in 2D building extraction is the image segmentation and classification technology. Known to all, the segmentation and classification is the fundamental technology for the thematic information extraction from remote sensing imagery, but also a long-standing bottleneck in digital image processing.

2.1 Image Segmentation and Classification Methods

As an overview, the recent research work on image segmentation and classification can be categorized into four classes, which are spectral-based (pixel-spectral), texture-based (gray-tone), object-oriented (region-growing), and spectral-spatial integration approach, see Figure 2-1.

Figure 2- 1. Multispectral segmentation and classification groups: spectral-based, texture-based, object-oriented, and spectral-spatial approaches.

2.2 Spectral-based Segmentation Algorithms

Spectral-based (pixel-spectral) approach focuses on image's multispectral properties, for example, pixel's RGB brightness values. It uses each pixel's spectral brightness values among various bands to partition an image, and each pixel is individually assigned to a specific class. It includes algorithms mainly based on parametric algorithm or nonparametric algorithm. Both parametric and nonparametric also belong to metric category, while a rule-based decision, like the tree classifier, is a nonmetric category algorithm.

Parametric classification algorithms, such as maximum likelihood classifier (MLC), assume normally distributed data and knowledge about the forms of the underlying class density functions (Duda et al., 2001). However, owing to the increased spectral multivariate or the internal variability, the accuracy of the MLC algorithm is low (Carleer et al., 2005). Consequently, it is not suitable for VHR satellite imagery.

Nonparametric classification algorithms, such as parallepiped, minimum distance to means, K-Means, Fuzzy C-Means (FCM) and ISODATA, may be applied to the data that are not normally distributed and without the assumption that the forms of the underlying densities are known (e.g., Friedl et al., 2002; Liu et al., 2002; Qui and Jensen, 2004). Unlike parametric methods, nonparametric methods are less sensitive to the increased spectral multivariate.

2.2.1 Minimum Distance to Means

As one of the earliest and basic spectral-based segmentation algorithm (Jahne, 1991; Lo et al., 2002), minimum distance to means classification algorithm belongs to nonparametric method, and is computationally simple and commonly used.

When used properly it can result in classification accuracy comparable to other more computationally intensive algorithms such as the MLC algorithm (Jensen, 2005). Like the parallelepiped classification algorithm, it requires that the user provide the mean vectors for each class in each band from the training data. To perform a minimum distance classification, a program must calculate the distance to each mean vector from each unknown pixel (Jahne, 1991; Yeung, 2002). It is possible to calculate this distance using Euclidean distance based on the Pythagorean theorem or “round the block” distance measures. When more than two bands are evaluated in a classification, it is possible to extend the logic of computing the distance between just two points in n space using the equation (Duda et al., 2001):

$$D_{AB} = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (2-1)$$

where D_{AB} is the distance between point A and B, n is the dimensions of a point, a_i , b_i are two point's coordinates, respectively.

By calculation the Euclidean distance from a point to the mean of all classes, it is possible to determine which distance is shortest, and the pixel point is assigned to the class from which the point obtains the minimum distance. It is obvious that any unknown pixel will definitely be assigned to one of the all classes.

2.2.2 K-Means

K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters). The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because different locations cause different results. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed. At this point we need to re-calculate k new centroids as barycentres of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop, the k centroids change their location step by step until no more changes are done.

Finally, this algorithm aims at minimizing an objective function, usually a squared error function.

The objective function

$$J = \sum_{j=1}^K \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (2-2)$$

where $\|x_i^{(j)} - c_j\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j , is an indicator of the distance of the n data points from their respective cluster centres.

Although it can be proved that the procedure will always terminate, the K-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. Overall, it does have some weaknesses:

- The way to initialize the means was not specified. It is significantly sensitive to the initial randomly selected cluster centres.
- The results produced depend on the initial values for the means, and it frequently happens that suboptimal partitions are found.
- It can happen that the set of samples closest to $\mathbf{m}_i$ is empty, so that $\mathbf{m}_i$ cannot be updated. This is an annoyance that must be handled in an implementation.
- The results depend on the metric used to measure $\| \mathbf{x} - \mathbf{m}_i \|$. A popular solution is to normalize each variable by its standard deviation, though this is not always desirable.
- The results depend on the value of k .

2.2.3 Fuzzy C-Means

Fuzzy C-Means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters to some degree that is specified by a membership grade. This method, first developed by Dunn (1973) and then improved by Bezdek in (1981), is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty \quad (2-3)$$

where m is any real number greater than 1 (note that the parameter m is case sensitive. When $m \rightarrow 1$, the fuzzy c-means converges to a “generalized” classical K-means; When $m \rightarrow \infty$, all clusters tend towards the centroid of data set X . that is, the partition becomes fuzzier with increasing m). u_{ij} is the degree of membership of x_i in the cluster j , x_i is the i th of d -dimensional measured data, c_j is the d -dimension centre of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the centre.

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centres c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}, \quad c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m} \quad (2-4)$$

This iteration will stop when $\max_{ij} \left\{ |u_{ij}^{(\kappa+1)} - u_{ij}^{(\kappa)}| \right\} < \varepsilon$, where ε is a termination criterion

between 0 and 1, whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of J_m .

Compared with K-means, data are bound to each cluster by means of a Membership Function, which represents the fuzzy behavior of this algorithm. To demonstrate that, an appropriate matrix named U could be built whose factors are numbers between 0 and 1, and represent the degree of membership between data and centres of clusters. Now, the following shows matrix U whose factors are the ones taken from K-means and the membership functions:

$$\begin{array}{cc}
 U_{KC} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ \dots & \dots \\ 0 & 1 \end{bmatrix} & U_{FCM} = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \\ 0.6 & 0.4 \\ \dots & \dots \\ 0.9 & 0.1 \end{bmatrix} \\
 \text{(a)} & \text{(b)}
 \end{array} \tag{2-5}$$

In the examples above we have considered the K-means (a) and FCM (b) cases. It is noticeable that in the case (a) the coefficients are always unitary, this is to indicate that each datum can belong only to one cluster, while (b) may not. That is the prominent advantage of FCM.

2.2.4 ISODATA

ISODATA stands for iterative self-organizing data analysis technique. This is a more

sophisticated algorithm that allows the number of clusters to be automatically adjusted during the iteration by merging similar clusters and splitting clusters with large standard deviations. It is similar to the K-means algorithm with the distinct difference that the ISODATA algorithm allows for different number of clusters while the K-means assumes the number of clusters is a priori.

The ISODATA algorithm is iterative because it makes a large number of passes through the remote sensing dataset until specified results are obtained, instead of just two passes. The iterative ISODATA algorithm is relatively slow, and image analysts are notoriously impatient. Analysts must allow the ISODATA algorithm to iterate enough times to generate meaningful mean vectors (Jensen, 2005). The ISODATA classification is generally performed through the following steps:

Step 1. Arbitrarily choose K initial cluster centres from the data set.

Step 2. Assign each of the samples to the closest cluster centre.

Step 3. Discard clusters with fewer than a threshold.

Step 4. Update each cluster centre.

Step 5. Compute the average distance of samples in a cluster from their corresponding cluster centres.

Step 6. Compute the overall average distance of the samples from their respective cluster centres.

Step 7. If there are too few clusters, go to Step 9; if there are too many clusters, go to Step 8; otherwise go to Step 10.

Step 8. Steps to split. Find the standard deviation vector for each cluster.

Step 9. Steps to merge.

Step 10. Terminate if maximum number of iterations I is reached. Recycle from Step 2.

Compared with K-means, both of these algorithms are iterative procedures. In general, both of them assign first an arbitrary initial cluster vector. The second stage classifies each pixel to the closest cluster. In the third stage, the new mean cluster vectors are calculated based on all the pixels in one cluster. The second and third stages are repeated until the "change" between the iteration is small. The "change" can be defined in several different ways: either by measuring the distances the mean cluster vector has changed from one iteration to another or by the percentage of pixels that have changed between iterations.

However, ISODATA algorithm has some further refinements by splitting and merging of clusters (Jensen, 2005). Clusters are merged if either the number of members (pixel) in a cluster is less than a certain threshold or if the centres of two clusters are closer than a certain threshold. Clusters are split into two different clusters if the cluster standard deviation exceeds a predefined value and the number of members (pixels) is twice the threshold for the minimum number of members. The ISODATA algorithm is more flexible than the K-mean method. But the user has

to choose empirically many more parameters listed previously.

2.2.5 Drawbacks of Spectral-based Algorithms

A general disadvantage of the K-means algorithm is that they work best for images with clusters that are spherical and have the same variance. Similarly, this is also true for ISODATA and fuzzy c-means algorithms. This is often not the case for the remote sensing images. For example, a cluster with "desert" pixels is compact (circular). A "forest" cluster, however, is usually more or less elongated (oval) with a much larger variability compared to the "desert" cluster. While the "desert" cluster is usually properly detected by the above algorithms as one distinct cluster, the "forest" cluster is often split up into several smaller clusters. In particular, the way the "forest" cluster is split up can vary quite a bit for different starting values in k-means and is thus arbitrary. Consequently, with the above algorithms, an extracted building roof is sometimes annoyingly split into many different roofs.

Another frustrating problem is the uneven illumination or reflection that can cause confusing results. For example, the sunlit and shady sides of the same roof have vastly different spectral responses, even though they belong to the same class (Thomas et al., 2003). Moreover, in spectral-based clustering, the number of clusters cannot usually be obtained directly and automatically.

Urban areas are much more heterogeneous than natural landscapes. With relatively low spectral resolution, the VHR satellite images have limited spectral capabilities (Carleer et al., 2005). Thus, the above multispectral per-pixel classifications from VHR satellite images constrict their effectiveness.

2.3 Texture-based Segmentation Algorithms

With the development of high spatial resolution imagery since the 1990s, for features that cannot be described easily, such as more complicated and natural image scenes, texture information is involved for segmentation from multispectral and hyperspectral sensor data. There is a large amount of literature on image segmentation based on the texture (Reed et al., 1993; Cheng et al., 2001; Zhang and Tan, 2002; Munnoz et al., 2003, Gong et al., 1990; Sharma and Sarkar, 1998).

In recent years, many algorithms are focusing on Markov Random Fields (MRF) using VHR satellite imagery or Synthetic Aperture Radar (SAR) imagery, such as hidden Markov models (Tso and Olsen, 2005), Gauss-Markov random field, Markov random field (MRF) clustering (Tran et al., 2005), Conditional Markov field (CMF) model and Tree Structure Markov random field (TS-MRF) (Poggi et al., 2005; Delia et al., 2003; Poggi et al., 1999). Overall, they are

grouped into structural, statistical and signal theoretic methods.

Structural methods are based on a more or less deterministic arrangement of textural elements. Statistical methods define colour textures as stochastic processes and characterize them by a few statistical features. Most relevant statistical approaches are co-occurrence matrices (Palm, 2004), semivariance, autocorrelation (Warner and Steinmaus, 2005; Puissant et al., 2005), and histograms (Paschos and Petrou, 2003). Signal theoretic approaches focus on periodic patterns resulting in peaks in the spatial frequency domain, e.g., Gabor filtering and wavelet decomposition (van de Wouwer et al., 1999; Xu et al., 2005).

As one of the typical gray-level texture algorithms, Markov Random Field segmentation algorithm will be reviewed in the following sections.

2.3.1 Markov Random Field and Model-based Segmentation

Markov random field (MRF) can be combined with model-based segmentation or clustering to take into account the spatial relation between pixels. In literature, the MRF model on model-based clustering first has been applied for the restoration of dirty images and referred to a smoothing technique that gives more weight to the fuzzy class memberships of spatial neighbor clusters. It is assumed that the class probability of a pixel is only dependent on class

memberships of its (spatial) neighbor clusters, so that it reduces the possible influence of noise and overlapping clusters (Tran et al., 2005).

More precisely, the w th-order neighborhood system for a particular pixel x_i , called ∂_i , is defined as a set of neighbor pixels belonging to a rectangular window of size w , centred at the pixel x_i . The conditional probability of pixel x_i of belonging to cluster c under the neighboring system ∂_i is estimated by (Besag, 1986)

$$P(c_i = c) = \frac{1}{Z} \exp \left[\beta \sum_{j \in \partial_i} u_{jc} \right] \quad (2-6)$$

where Z is a normalization constant and β is a spatial smoothness parameter. A higher (positive) β corresponds to higher spatial dependency of neighbor pixels. The EM algorithm is then adapted, leading to the log likelihood criterion

$$\log L_{MRF}(\Psi) = \sum_{c=1}^g \sum_i^N u_{ic} \log(\pi_{ic} f(x_i; \theta_c)) \quad (2-7)$$

The mixture proportions π_c are now replaced by the transition probability π_{ic} (McLachlan et al., 2000)

$$\pi_{ic} = \frac{\exp \left(\beta \sum_{j \in \partial_i} u_{jc} \right)}{\sum_{h=1}^g \exp \left(\beta \sum_{j \in \partial_i} u_{jh} \right)} \quad (2-8)$$

Not only does the algorithm tend to converge faster, since EM tends to converge to a local optimum, the segmentation accuracy is heavily dependent on the initial guess Ψ_o and the

choice of number of clusters (Besag, 1986; Fjortoft et al., 2003). Therefore, obtaining a good Ψ_o is a key but also a challenge element of MRF segmentation.

2.3.2 Discussion

Generally, the results obtained from texture approaches are barely satisfactory for colour images; some of them gained better results in hyperspectral, but not for VHR satellite imagery. When considering the texture features, almost all the images are processed in gray tone, i.e., the spatial texture information was extracted independently in single panchromatic band or gray level imagery, with little interactive relation to multispectral information.

First of all, it is important to note that for complex land surfaces, particularly when sensed using high-resolution multi-spectral satellite imagery, the image scene is usually not easily partitioned correctly using a single feature channel (property) due to the complicated nature of the image scene. Integrated processing of the multiple features is essential to obtaining more reliable segmentation (Hu et al., 2005). Furthermore, in remote sensing images, information is recorded from different parts of the spectrum. Therefore, textures in these bands are not necessarily similar. In between band relations should be taken into account (Lucieer et al., 2005).

2.4 Object-oriented Segmentation Algorithms

Contrary to pixel-based spectral methods, another major category of image segmentation and classification is the object-oriented (region-growing) approach. It utilizes the statistics of region homogeneity and so tends to be less sensitive to noise (Turner et al., 2001; Herold et al., 2002; Batistella et al., 2003; Gorte and Stein 1998; Lucieer and Stein, 2002; Benz et al., 2005; Wong et al., 2005; Heene et al., 2000; Mansor et al., 2005; Forman, 1995).

The concept behind object-oriented segmentation is that important semantic information is not represented in single pixels, but in meaningful image objects and their mutual relationships. The formation of the objects is carried out in a way that an overall homogeneous resolution is kept. The object-oriented algorithm process in the eCognition software can be split into two steps, segmentation and classification. Objects were assigned class rules using spectral signatures, shape and contextual relationships. The rules were then used as a basis for the fuzzy classification (Whiteside and Ahmad, 2005), see Figure 2-2.

Figure 2- 2. Schematic diagram illustrates the classification flow in eCognition object-oriented approach (Wong et al., 2005).

Scale is a crucial aspect of image understanding. It describes the magnitude or the level of aggregation on which a certain phenomenon can be described. The same type of object appears differently at different scales. Vice versa, the classification task and the respective objects of interest directly determine a particular scale of interest.

Figure 2-3 shows the object-oriented segmentation with the scale concept. All of the image objects were automatically linked to a network after the segmentation process. Subsequently, different scale parameter creates a hierarchical network of image objects. Each image object knows its super-object and sub-objects. The basic difference, especially when compared to pixel-based procedures, is that object oriented analysis does not classify single pixels, but rather image objects which are extracted in a previous image segmentation step.

Figure 2- 3. The concept of scale, in which mainly three different levels of image objects have been created representing different scales (Wong et al., 2005).

However, scale is also a weakness of object-oriented segmentation. Although the object-oriented approach is generally better than texture segmentation or spectral-pixel based approaches in land-cover images and apparently deletes the salt-and-pepper effect; scale brings a lot of

limitations in its applications. Most of all, the object-oriented approach is not suitable for building extraction because it is easy to create problems in defining the segmentation scale level for a particular object, and also difficult to assign correct weights for multi-sized building objects during the segmentation process, i.e., missing of buildings after classification process is a problem (Mayunga, 2006). Moreover, the uncertainty of segmentation output in the building roof outlines (transition zones) is still unsatisfactory (Whiteside and Ahmad, 2005; Carleer et al., 2005).

2.5 Spectral-spatial Segmentation Algorithms

People came to realize that for complex land surfaces, particularly when sensed using multi-spectral high-resolution satellite imagery, the image scene is usually not easily partitioned correctly using a single feature channel (property) due to the complicated nature of the image scene. Integrated processing of the multiple features is essential to obtaining more reliable segmentation (Hu et al., 2005). On the one hand, taking into account only spectral properties causes problems because identical spectral reflectance values can correspond to very different objects (for instance, a concrete flat roof can reflect as a road). On the other hand, taking into account only geometric criteria also causes problems because a similar geometric shape can refer to very different objects. The extraction of spatial information in addition to spectral properties is then often used (Gong et al., 1992; Sharma and Sarkar, 1998).

Several authors have demonstrated that spatial and spectral combined information can lead to significant improvements in the classification accuracy of urban or built-up areas. The output image generated by texture analysis is often classified directly or used as an additional band together with other multi-spectral bands in classification (Gong and Howarth, 1990; Marceau et al., 1990). Unfortunately, these procedures have been tested on medium resolution satellite images (e.g., 10-20 m SPOT-4 imagery), but they have not been applied on VHR satellite imagery (Puissant et al., 2005).

Spectral and texture analysis in classification are two basic tasks in building extraction. The dependencies of both are still not clear and, hence, several concepts dealing with coloured textures exist (Palm, 2004), and these hybrid concepts are generally called the spectral and spatial integration approach. There are three typical strategies for the spectral-spatial integration approach; namely, spectral-spatial sequential integration, weighted additive combination, and integrative colour-texture vectors.

2.5.1 Spectral-spatial Sequential Integration

Spectral-spatial sequential combination is the algorithm that takes into account both the spectral and spatial information in a sequential segmentation or classification process. It usually has less operation complexity, high degree of accuracy, computer efficiency, and can work on natural

colour images directly. In particular, so far, it may be the best segmentation or classification approach for building extraction from VHR satellite imagery.

Owing to its reliability, there are a great number of researches or applications on it (Kasari et al., 1999). For instance, Song et al. (1996) proposed an algorithm that uses colour and texture information to detect the defect of granite with random texture.

In this algorithm, sequential chromatic-structural approach was proposed, which combines colour clustering with a binary blob image analysis to capture the relevant information content of the textures. Figure 2-4 illustrates that this colour texture defect detection algorithm consists essentially of two stages: the first stage is training and colour classification. The second stage is structural feature extraction and blob classification. The final step is analyzing the output granite image for the presence of any defects, as well as detecting their locations in the image.

Figure 2- 4. Flowchart of the granite defect detection

Accurate results were obtained (see Figure 2-5) in the structural feature extraction and blob classification with many sub-clusters, whereas spectral classification only used two main classes. As a result, by segregating the colour image texture into smaller texture clusters, better class representation was obtained.

Figure 2- 5. (a) Data are poorly represented by clusters A and B; (b) super clusters A and B are formed by many sub-clusters. (Song et al., 1996)

The most typical research of this sequential integration is the Extraction and Classification for Homogenous Objects (ECHO) (Landgrebe, 1980; Mirmehdi and Petrou, 2000) or unsupervised ECHO called UnECHO (Jimenez et al., 2005). In common, they sequentially combine spectral and spatial information by analyzing spatial interactions of colours in a given neighborhood, and use the concept of degree of membership of a neighborhood to a single cluster. This degree of membership verifies the degree of heterogeneity, or homogeneity, of a particular neighborhood.

UnECHO has two stages. The first stage consists of a conventional clustering algorithm that classifies the image on a pixel-by-pixel basis, i.e., C-means. The second stage uses both, the

classification results of the first stage and the spectral content in the image.

UnECHO divides the whole multispectral or hyperspectral image into a number of non-overlapping neighborhoods. The algorithm will measure the degree of heterogeneity of all neighborhoods. If the k th is spectrally homogeneous, which implies low heterogeneity, then all pixels in this neighborhood $\hat{X}^{(\kappa)}$ will be forced to be the i th cluster according to the following rule.

$$\text{if } Q_i(\hat{X}^{(\kappa)}) < T_i \text{ and } Q_i(\hat{X}^{(\kappa)}) < Q_j(\hat{X}^{(\kappa)}) \quad \forall j \neq i \quad (2-9)$$

The first proposition, $Q_i(\hat{X}^{(\kappa)}) < T_i$, verifies if the degree of homogeneity of the κ th neighborhood is large enough. If yes, then that particular neighborhood can be classified as a single cluster according to the second proposition. If that is not the case, $\hat{X}^{(\kappa)}$ is spectrally heterogeneous and has a high degree of membership deficiency to a single cluster. On this condition, the algorithm does not change the classification performed on a pixel-by-pixel basis for that particular neighborhood that resulted from the first stage.

The second proposition, $Q_i(\hat{X}^{(\kappa)}) < Q_j(\hat{X}^{(\kappa)})$, assigns all the pixels in the neighborhood $\hat{X}^{(\kappa)}$ to the i th cluster if the degree of membership (degree of homogeneity) to that particular i th cluster is the largest (smallest $Q_i(\hat{X}^{(\kappa)})$). An advantage of this mechanism is that it is fast and integrates at the same time the spatial and the spectral information. $Q_i(\hat{X}^{(\kappa)})$, which is the degree of heterogeneity and can be understood as a degree of membership deficiency, is inversely

proportional to the degree of homogeneity and cluster membership. It is defined as the average of the distances of every pixel to the j th cluster

$$Q_j(\hat{X}^{(k)}) = \frac{1}{L} \sum_{m=1}^L q_j(X_m) \quad (2-10)$$

Where X_m is the m th pixel in the k th neighborhood $\hat{X}^{(k)}$ and L is the number of pixels in that neighborhood. $q_j(X_m)$ is the distance of the pixel to the j th cluster.

In an experiment of this approach, the hyperspectral data used is a segment of a HYDICE frame. HYDICE is a hyperspectral sensor that collects data in 210 bands with an IFOV of 3.5 m. Figure 2-6 (Jimenez et al., 2005) shows the segment of the analyzed frame. This segment has two well known landmarks: the roof of a supermarket, in white colour, and the asphalt of a parking lot below the white structure. Figure 2-7a (Jimenez et al., 2005) shows the results of the C-means clustering used in the first stage of the algorithm and the results of the second stage in UnECHO (see Figure 2-7b) (Jimenez et al., 2005) that integrate the spatial and the spectral information.

Figure 2- 6. Segmentation of a HYDICE hyperspectral image (Jimenez et al., 2005)

The result of UnECHO classifier, shown in Figure 2-7b, possesses more details and is better defined in terms of the spatial structures such as buildings and parking lots. For example, UnECHO classified as one cluster, red colour, and the roof of a supermarket. Meanwhile in Figure2-7a, pixel-by-pixel C-means clustering, it appears as two different and not well-defined clusters. In Figure 2-7b, the asphalt in the whole image is mostly contained in one cluster, blue colour, facilitating the process of identifying a parking lot and the roads in the class map. In Figure 2-7a, the asphalt appears in at least two distinct clusters, blue and green, making more difficult the identification process after applying an unsupervised classification algorithm.

Figure 2- 7. (a) Result of C-means algorithm applied at first stage, (b) Result of UnECHO algorithm applied at the second stage. (Jimenez et al., 2005)

According to the data analysis of these results, UnECHO enhanced the spectral-based classification accuracy of C-means by a total of 16%. UnECHO finds more objects with more detailed spatial structures that are meaningful to a human analyst and facilitates the identification process for spatial structures in an urban area, e.g., the orange and yellow clusters. Meanwhile, in pixel-by-pixel classification, due to the speckle appearance, there is a lack of objects that could have spatial meanings for a human analyst, like buildings, roads, and cars. In general, UnECHO presents a more clear class map, enhancing borders of regions with different shapes, without losing the particular details of small objects like cars and narrow lines.

2.5.2 Weighted Additive Combination

This is a type of additive combination of multiple image features, including texture, intensity and colour information. In the earlier research, each colour band is processed separately, and then the researchers simply extended gray-level texture analysis to colour images (Caelli and Reye, 1996; Eom, 1999). Chen and Chen (2002) measured the homogeneity or similarity of the colour and texture feature with an aggregate function of the two features in which the weights of each feature are set as constants. While considering the complexity of high-resolution satellite images, Hu et al. (2005) made the weights of each feature more adaptive for feature combination. Their region-based segmentation is a typical example of the weighted additive combination approach and their algorithmic framework consists of three steps: (1) hierarchical splitting; (2) agglomerative merging, and (3) pixel-wise refinement (see Figure 2-8) (Hu et al., 2005). In the hierarchical splitting, different regions are separated according to texture, intensity, etc. The second step merges similar adjacent sub regions until a stopping criterion is triggered. The segmentation is completed using pixel-wise refinement in order to improve localization of the region boundaries.

(a)

(b)

Figure 2- 8. Region-based segmentation; split-merge plus pixel-wise refinement: (a) the workflow, (b) region similarity measure is critical. (Hu et al., 2005)

Hu et al. (2005) used feature distribution to describe the features. The estimation of the joint distribution of multi-channel features is to compute similarity in each feature space separately and then to integrate the scores into an aggregate similarity score. To measure the similarity

$\text{Sim}_{ij}(f_t, f_i, f_c)$ of two regions i and j , an aggregate function to combine multiple features was used:

$$\text{Sim}_{ij}(f_t, f_i, f_c) = C_t \rho_t + C_i \rho_i + C_c \rho_c \quad (2-11)$$

$$C_t = V_t / (V_t + V_i + V_c) \quad (2-12)$$

$$C_i = V_i / (V_t + V_i + V_c) \quad (2-13)$$

$$C_c = V_c / (V_t + V_i + V_c) \quad (2-14)$$

where (f_t, f_i, f_c) represents the three different types of features, which are described as the discrete histograms: (ρ_t, ρ_i, ρ_c) are respectively similarity measures of texture, intensity, and colour features. Now, the problem is how to adaptively determine the weights of the features rather than simply using the constant weights. Hu et al. (2005) used adaptive determination to derive the texture, intensity, and colour weights (C_t, C_i, C_c). The adaptive estimation of the weights of the multiple features for similarity measure of two regions i and j :

$$V_t = \max(p_{ti}, p_{tj}) * (1 - \rho_t) / 0.2 \quad (2-15)$$

$$V_i = [1 - \min(p_{ti}, p_{tj})] * (1 - \rho_i) * |\bar{g}_i - \bar{g}_j| / 128 \quad (2-16)$$

$$V_c = \max(p_{ci}, p_{cj}) * (1 - \rho_c) * (1 - V_t) * (1 - V_i) \quad (2-17)$$

where (V_t, V_i, V_c) are the weight estimation of the texture, intensity, and colour features; (p_{ti}, p_{tj}) and (p_{ci}, p_{cj}) are the estimation of region i and j being a texture and colour region; and $\bar{g}_i$ and $\bar{g}_j$ are the mean values of the intensity of the two regions.

Overall, the adaptive integration of multiple features provides more reliable and accurate segmentation results than when using just a single feature. Nevertheless, the thresholds have

considerable impact on segmentation results. For example, the initial divided window size, the level number of splitting and the trigger criterion x are essential, also, the stopping criterion affects the segmentation result. Especially for small objects, like various sized buildings, the possible means for selecting the proper thresholds are difficult and limited.

2.5.3 Integrative Colour-texture Vectors

Integrative colour-texture vectors is the algorithm which takes into account cross-band relations between pixels, treating colour and texture as a joint phenomenon (joint vectors) instead of considering texture (spatial distribution) only in one spectral band (gray-tone) or in colour bands individually.

Currently, there are four kinds of main branches, one of which is based on the development of Local Binary Pattern (LBP) (e.g., Ojala and Pietikainen, 1999; Lucieer et al., 2005). Lucieer et al. (2005) proposed a multivariate extension of the standard univariate LBP operator to describe integrative colour texture. The second is based on Gabor filter (Hoang et al., 2005; Thai and Healey, 1998). The third is based on the Co-occurrence Matrices Algorithms, for example, Colour Texture Classification by Integrative Co-Occurrence Matrices in LUV Colour Space (Palm, 2004), Colour Texture Analysis for Quantitative Laryngoscopy (Ilgnier et al., 2003). The fourth is based on the development of Markov Random Field (Xu et al., 2005; Suen and Healey,

1999; Panjwani and Healey, 1995).

However, these methods often transform colour images into gray-level images in terms of intensity values (Ojala and Pietikainen, 1999), edge values (Shafarenko et al., 1997) or quantization labels (Chang and Wang, 1996). Often, MRF-based unsupervised classification algorithms are highly complex, taking long periods of data processing (Jimenez et al., 2005). In common, these methods could not work directly on the colour image. Their generalization is not always straightforward, usually applied to hyper-spectral imagery, and has many limitations or is case sensitive in natural colour images. As for the VHR satellite imagery, they are still in conceptional application.

2.6 Summary of Existing Methods

The typical algorithms, types, advantages, and limitations of various algorithms are summarized for each group in Table 2-1. It lists major classes' categories and their typical algorithms, type, and finally, the comparison of their advantages and drawbacks. Based on the following summary, Spectral-spatial integration classification demonstrated its superiority over others, and it has the most promising future.

Table 2- 1.The summary of existing methodologies of multispectral imagery segmentation

Methods	Typical Algorithms	Type	Advantages	Limitations
Spectral based	Tree-classifier, Maximum Likelihood Classifier, Minimum Distance to Means, K-means, FCM, ISODATA	Parametric, non-parametric, hard, soft, nonmetric, hybrid	<ul style="list-style-type: none"> ♦ Good use of spectral information ♦ Simple algorithms ♦ Computational effective 	<ul style="list-style-type: none"> ♦ Low accuracy ♦ Lack of texture information ♦ Not suitable for VHR satellite image
Texture based	MRFs series, Co-occurrence Matrices, Semivariance, Autocorrelation, Gabor, Wavelet	Structural Statistical Signal	<ul style="list-style-type: none"> ♦ Good use of spatial information 	<ul style="list-style-type: none"> ♦ Applied to gray-tone, not natural colour or VHR satellite image; ♦ Sensitive to noise ♦ Computational consuming
Object oriented	eCognition texture descriptor, LBP/C texture descriptor	Hybrid	<ul style="list-style-type: none"> ♦ Higher accuracy ♦ Good use of spatial homogeneity 	<ul style="list-style-type: none"> ♦ Difficult scale selection, not suitable for building extraction ♦ Mainly for gray-tone ♦ High uncertainty on boundaries

Spectral-spatial integration	Spectral-spatial sequential integration		<ul style="list-style-type: none"> ♦ High accuracy ♦ Good use of both spatial and spectral information ♦ Can work both on natural colour and VHR satellite imagery ♦ Less operation complexity and computer efficiency 	<ul style="list-style-type: none"> ♦ Sensitive to neighborhood size
	Weighted additive combination		<ul style="list-style-type: none"> ♦ High accuracy ♦ Good use of texture and spectral information ♦ Adaptive weights for feature combination ♦ Applied to VHR satellite imagery 	<ul style="list-style-type: none"> ♦ Difficult threshold selection ♦ More suitable for land-cover extraction instead of building extraction
	Integrative colour-texture vectors		<ul style="list-style-type: none"> ♦ Much better use texture information than spectral information 	<ul style="list-style-type: none"> ♦ Highly complex ♦ Best applied to hyper-spectral imagery instead of VHR satellite imagery ♦ Can not work directly on the colour images ♦ Computational consuming

Chapter 3

PROPOSED METHODOLOGY

In this chapter, an algorithm is proposed towards automated extraction of 2D building from IKONOS imagery by integrating spectral and spatial information. In this chapter, some theoretical background is given first and then the framework and flowchart of proposed methodology are explained. Following that is the detailed description of proposed methodology.

3.1 Introduction

3.1.1 RGB Colour Space

A colour space is a means of uniquely specifying a colour. For example as humans we normally determine colour by such parameters as brightness, hue, and colourfulness. In computers it is more common to describe three colour guns of red, green and blue (RGB). These are related to the excitation of red, green, and blue phosphors on a computer monitor.

As one of various colour spaces, the RGB colour space is a projection of the cumulated intensity of reflected light within a defined range of wavelength to the axis of a 3D space. Because each

colour in RGB colour space only has the discrete values between zero and some maximum intensity (255 for 8-bit length), the resulting structure of this colour space is a cube. Any colour can be simply defined by giving its red, green, and blue values or coordinates within the colour cube (Pratt, 1991). These coordinates are usually represented as an ordered triplet (r, g, b) . An example of such a cube is shown below (Figure 3-1).

Figure 3- 1. The RGB colour space and cube.

Because of the disproportional increase of feature space's dimensionality, the best results in RGB are achieved with the parallel concept colour analysis by feature distribution (histogram).

3.1.2 3D Colour Feature Distribution (Histogram)

Colour histogram is an important tool in colour image analysis, because of its efficiency, effectiveness and triviality in computation (Pratt, 1991). Generally speaking, a colour histogram represents the statistical distribution of the colours in a colour image on all colours in a colour space. From the view of the applicability and computation, the common way is to divide a colour space into many bins (i.e. the discrete values intervals between zero and the maximum intensity value) and calculate the colour distribution of those bins.

Given a colour space divided into I colour bins, the colour histogram of the colour image CI with n pixels is represented as a vector $H = [h_0, h_1, \dots, h_{I-1}]$, in which each entry h_i indicates the statistical frequency figures of the colours in the colour image which belong to the i th bin, i.e.,

$$h_i = \frac{n_i}{n}, i = 0, 1, \dots, I-1 \quad (3-1)$$

where n_i is the number of pixels with colours in the i th colour bin.

In this thesis, RGB colour space is represented with 24 bits (8 bits for each of the three

components). Each colour corresponds to a point in RGB colour space. In order to obtain a colour histogram, pixels of a colour image distributed on all bins should be accounted for and shown by expanding them in a linear axis.

Let RGB colour space be discretized along the R, G, and B axis by the numbers N_R , N_G , and N_B , respectively. Then the total $N (= N_R \times N_G \times N_B)$ bins are available. These bins are coded in such a sequence from R to G and then from G to B. According to the specified discretizing and coding scheme the index of each bin can be represented as

$$i = R + N_G \times G + N_B^2 \times B \quad (3-2)$$

where $R = 0, 1, \dots, N_R-1$, $G = 0, 1, \dots, N_G-1$, and $B = 0, 1, \dots, N_B-1$.

Then the pixel (r_p, g_p, b_p) will be in the bin with the index i_p ,

$$i_p = \left\lfloor \frac{r_p N_R}{256} \right\rfloor + N_G \times \left\lfloor \frac{g_p N_G}{256} \right\rfloor + N_B^2 \times \left\lfloor \frac{b_p N_B}{256} \right\rfloor \quad (3-3)$$

where the floor function $\lfloor \rfloor$ is an integral operator.

Figure 3-2. shows an example of the divided RGB colour space with $N_R = N_G = N_B = 4$ and the

codes of all bins. In this example, a total of 64 ($=4 \times 4 \times 4$) bins can be obtained.

Figure 3- 2. Discretization of RGB colour space and codes of bins.

Figure 3-3 shows an example of a colour image's (right top corner) feature distribution diagram with total 512 ($=8 \times 8 \times 8$) bins.

Figure 3- 3. A 3D colour feature distribution (histogram) with 512 bins
(the green dots represent the major distribution peaks).

3.1.3 Morphological Operation

The morphological operation (MatLAB v6.5, 2004) is successful at removing much of the structural clutter, and performs well in comparison with traditional majority filtering to reduce

local spectral variation. There are various common morphological operation tasks with different effects and functions in post-image processing, such as dilation and erosion, contrast enhancement, noise removal, thinning, filling, and segmentation. In this thesis, the morphological operation topic is mainly focused on dilation and erosion.

Dilation and erosion are the two fundamental morphological operations. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries. After dilation and erosion the speckle pixels or sporadic outlier pixels in an image will be removed and the major objects will become more clear and prominent.

In the morphological dilation and erosion operations, the state of any given pixel in the output image is determined by applying a rule to the corresponding pixel and its neighbors in the input image. The rule used to process the pixels defines the operation as a dilation or an erosion. Table 3-1 lists the rules for both dilation and erosion.

Table 3- 1. Rules for grayscale dilation and erosion

Operation	Rule
Dilation	The value of the output pixel is the maximum value of all the pixels in the input pixel's neighborhood. In a binary image, if any of the pixels is set to the value 1, the output pixel is set to 1.

Erosion	The value of the output pixel is the minimum value of all the pixels in the input pixel's neighborhood. In a binary image, if any of the pixels is set to 0, the output pixel is set to 0.
---------	--

Figure 3-4 illustrates the dilation of a binary image in which the structuring element defines the neighborhood of the pixel of interest that is circled. The dilation function applies the appropriate rule to the pixels in the neighborhood and assigns a value to the corresponding pixel in the output image. In Figure 3-4, the morphological dilation function sets the value of the output pixel to 1 because one of the elements in the neighborhood defined by the structuring element is one.

Figure 3- 4. Morphological Processing of a Binary Image

Dilation and erosion are combined to implement image processing operations. MatLab is used for image processing. The definition of a morphological ‘opening’ of an image is an erosion followed by a dilation. The related operation, morphological ‘closing’ of an image is the reverse: it consists of dilation followed by erosion with the same structuring element. The following example uses the combination of dilation and erosion (MatLab functions: ‘imdilate’ and ‘imerode’) to illustrate how to implement a morphological opening or closing. Figure 3-5 shows the original building image and its roof extraction in binary image. Figure 3-6 shows the corresponding results during the ‘opening’ morphological operation, while Figure 3-7 shows the corresponding results during the ‘closing’ morphological operation.

Figure 3- 5. Original building image (left); and building roof extracted in binary image (right).

Figure 3- 6. The ‘opening’ morphological operation: erosion (left) + dilation (right).

Figure 3- 7. The ‘closing’ morphological operation: dilation (left) + erosion (right).

From the above morphological operation combination, the ‘closing’ combination of morphological operation is more suitable for building roof extraction by its more accurate results.

3.2 Flowchart of Proposed Methodology

For the task of extracting 2D buildings from IKONOS imagery, this proposed approach is based on a spectral-spatial sequential integration algorithm. Figure 3-8 illustrates the flowchart of the proposed approach.

Figure 3- 8. Flowchart of spectral-spatial integration.

Starting with the Multi-peak Supervised Segmentation (MSS) on 3D colour feature distribution, this automated spectral-spatial integration extraction approach generates spectral-based classification first, and transfers it into binary image. Again, with the MSS, it also generates a spatial-based structure classifier for each pixel neighborhood. Eventually, the original colour image and the result of spectral-based classification are fed into the spatial structure classifier for further spectral-spatial integration classification.

According to the spatial analysis of the neighborhood's degree of membership to classes, homogeneity and cross-neighborhood structure, neighborhood groups are to be judged whether or not the previous classification should be revised in each neighborhood group. In this major step, the spectral and spatial information are interactively integrated, and the output is a substantially improved classification due to the above rationale.

Next process is the post-processing. It includes the 'closing' morphological operation, and the filtering of small irrelevant objects, which filter impurities and keep the real important information for the final building extraction.

Finally, buildings and their boundaries are extracted; moreover, for the comparison with the original building image, the extracted building roof outlines are superposed on the original image^e for visual evaluation.

In this proposed approach, the Multi-peak Supervised Segmentation (MSS) is a new and effective segmentation algorithm owing to its easy signature sample selection and high degree of accuracy. Moreover, the spatial structure classifier is developed from the well known Extraction and Classification for Homogenous Objects (ECHO) classifier (Jimenez et al., 2005; Landgrebe, 1980; Richards, 1993). Two other new algorithms are also exploited in this spatial structure classifier: one is the Overlapped Neighborhood Analysis (ONA) instead of non-overlapped neighborhood structure in ECHO; another is the Cross-neighborhood Structure Detection (CSD).

3.3 Description of Algorithms

As introduced in Chapter 2, sequential spectral-spatial integration approach can work on natural colour images directly, has a high degree of accuracy and computer efficiency. So far, it may be the best segmentation or classification approach for building extraction from VHR satellite imagery.

However, traditional Extraction and Classification for Homogenous Objects (ECHO) (Landgrebe, 1980; Mirmehdi and Petrou, 2000) or unsupervised ECHO called UnECHO (Jimenez et al., 2005) has weaknesses in spectral segmentation, and non-overlapped neighborhood structure, especially the final product accuracy needs improvement.

In this thesis, a developed sequential spectral-spatial integration approach overcomes those weaknesses in ECHO or UnECHO, and it improved the Kappa statistics (κ) from 72.2% to 83.7% as shown in the experiments. These percentages show a significantly high accuracy based on conditions equivalent to those in current literature. The following sections will describe its aforementioned three innovative highlights: Multi-peak Supervised Segmentation (MSS), Overlapped Neighborhood Analysis (ONA) and the Cross-neighborhood Structure Detection (CSD).

3.3.1 Supervised Segmentation Logic

Supervised segmentation is the process of using priori samples of known identity (pixels already assigned to the information classed, or called feature signatures) to classify unknown identity (Besag, 1986). The analyst defines training areas by identifying regions on the image that are matched to areas of known identity on the ground. The selection of training areas is a key step in supervised segmentation.

Conventionally, in building extraction, the unsupervised segmentation was usually used. But when the surface reflectance of the object is not uniform from the background and the image scene is not evenly illuminated, the resulting unsupervised segmented image will be a mess; and if a building has multi-material or has similar colour and texture to its background, separating it

away from the environment is very difficult, even though the colour-texture combined approach is used. However for human eyes, there are no such problems; it is a piece of cake to tell multi-coloured, multi-material or uneven reflected buildings from their environment, why? Experience does the job! Nevertheless, supervised segmentation can give the descriptor good experience. Supervised segmentation uses priori samples of known identity (human experience) to classify unknown objects; therefore, it is a promising approach to solve the multivariate weakness in VHR imagery. But conventional supervised segmentation always has an annoying obstacle in priori sample (signature) selection: their final results are very sensitive to the chosen priori samples, which will result in strong bias in the final classification. Thus, an innovative supervised segmentation-Multi-peak Supervised Segmentation (MSS) is needed.

3.3.2 Multi-peak Supervised Segmentation

Conventional supervised training sample (feature signature) selections are very sensitive to the sample image's size, impurities, context change, and subjective selection. They usually calculate all the pixels' arithmetic average, weighted average, or variance-modified average in a training sample to extract the feature signature. The number of total pixels or the variance of the pixel value directly affects the feature signature and different people with different selection habits in an image also matters, as a result, strong biases are conveyed to the final segmentation.

Moreover, the post-processes to evaluate and revise these training samples are also cumbersome. First of all, they need a number of sample images for signature convergence and impurity filtering, and then they need to extract their stable signatures. Also, they use 'Alarm', 'Ellipse', 'Contingency matrix', 'Divergence', and 'Statistics' tools to verify and revise these signatures, and make sure of their effectiveness by judging the Separability, Weight Factors, Probability of Error and so on (ERDAS, 1999).

Nevertheless, biases were inevitable in the final segmentation results. This is a key issue and long-standing problem in supervised segmentation, and also explains why sometimes people prefer unsupervised segmentation to supervised segmentation.

To overcome this barrier, we need to pinpoint the real and most effective signatures in a feature sample. The MSS is exactly such a kind algorithm.

In MSS, a 3D colour histogram is used as an effective tool to capture the typical signatures hiding in a feature sample. For example, Figure 3-9 illustrates that in a random sized sample image-'BUSH', there are five hiding peaks which get more prominent high frequency than others. In Figure 3-10, a random sized sample image-'ROOF' has nine outstanding pixel bin peaks, whereas in Figure 3-11, the 'BARELAND' only owns four giant peaks. (Note that the selection of all these samples was not restricted by size, object direction, or shape.)

Figure 3- 9. 100×79 ‘BUSH’ image (top right) has 5 prominent peaks covering $\geq 70\%$ pixels.

Figure 3- 10. 37×36 ‘ROOF’ image (top right) has 9 prominent peaks, covering $\geq 60\%$ pixels.

Figure 3- 11. 56×24 pixels ‘BARELAND’ image (top right) has only 4 prominent peaks,
but covers $\geq 90\%$ pixels.

From the above examples, two common characteristics can be summarized as follows:

First, each image has its feature peaks, i.e., highly concentrated and apparently separated typical pixel groups, like genes. They cover the extreme majority and are not evenly distributed, moreover, separated, which means the traditional extraction algorithms of arithmetic average, weighted average, and variance-modified average are not appropriate. Take the data shown in

Figure 3-9 as an example, the top two peaks look like they are very close, almost joined together. Consequently, it seems reasonable to take their arithmetic average as the 'BUSH' signature. But it can be seen that the bin numbers of these top two peaks are 1124 and 1184 respectively, i.e., there is a difference of 60 bins between them. With 16 bins in each colour axis (i.e., each bin gets $256 \div 16 = 16$ spectral value in one dimension), the real distance between these two peaks is amazingly 960 brightness values. That is not a small crevice, but a 'gorge', considering the 30 or 50 are big enough brightness value for thresholds of the signature judgment in radial basis function algorithms. Obviously, previous simple averages ignored these 'gorges', which resulted in misleading results.

Second, these peaks are non-related to original sample image size, shape or direction, and easy to be selected in an image. As long as we select the spots that approximately include them, these 'genes' will be captured in the colour histogram owing to their highly concentrated distribution, regardless of many impurities. However, conventional sample selections have demanding requirements for image size, shape or purity in sample selection.

Like a precisely guided missile, MSS takes the above characteristics into account, and calculates the precise Euclidean distance between a targeted pixel (or a neighborhood) and every individual peak disclosed in a theme sample histogram. Then it finds the minimum distance that a targeted pixel (or a neighborhood) owns in a theme sample. After that, it further chooses the global

minimum distance among all the previous minimum distances found in all the theme samples and finally labels the theme sample from which the targeted pixel got the global minimum distance. Then, MSS assigns the pixel (or the neighborhood) to the theme class for which the labeled theme sample stands.

In a summary, MSS is a precisely guided segmentation algorithm, which overcomes the internal variability of a VHR sample image. It requires fewer samples (usually only 1-4 random-sized samples for a theme class) with simpler and quicker sample selection.

3.3.3 Integration of Spectral and Spatial Information

Extraction for Homogenous Objects

This approach sequentially combines spectral and spatial information by analyzing spatial interactions of colours in a given neighborhood, and uses the concept of degree of membership of a neighborhood to a single class. This degree of membership verifies the degree of heterogeneity, or homogeneity of a particular neighborhood.

It has two stages. The first stage consists of an innovated Multi-peak Supervised Segmentation (MSS) algorithm that classifies the image on a pixel-by-pixel basis. The second stage uses both

the classification results of the first stage and the original spectral content in the image.

It divides the whole multispectral or hyperspectral image into a number of overlapped neighborhoods. The algorithm will measure the degree of heterogeneity of all neighborhoods. If the k th is spectrally homogeneous, which implies low heterogeneity, then all pixels in this neighborhood $\hat{X}^{(k)}$ will be forced to be the i th class according to the following rule.

$$\text{if } Q_i(\hat{X}^{(k)}) < T_i \text{ and } Q_i(\hat{X}^{(k)}) < Q_j(\hat{X}^{(k)}) \quad \forall j \neq i \quad (3-4)$$

The first proposition, $Q_i(\hat{X}^{(k)}) < T_i$, verifies if the degree of homogeneity of the k th neighborhood is large enough. If yes, then that particular neighborhood can be classified as a single class according to the second proposition. If that is not the case, $\hat{X}^{(k)}$ is spectrally heterogeneous and has a high degree of membership deficiency to a single class. On this condition, the algorithm does not change the classification performed on a pixel-by-pixel basis for that particular neighborhood that resulted from the first stage.

The second proposition, $Q_i(\hat{X}^{(k)}) < Q_j(\hat{X}^{(k)})$, assigns all the pixels in the neighborhood $\hat{X}^{(k)}$ to the i th class if the degree of membership (degree of homogeneity) to that particular i th class is the largest (smallest $Q_i(\hat{X}^{(k)})$). That means, if the average distance of all the pixels in neighborhood $\hat{X}^{(k)}$ to the i th class is the shortest compared with any other class, then this

neighborhood $\hat{X}^{(k)}$ has the highest tendency of belonging to the i th class.

An advantage of this mechanism is that it is fast and it integrates at the same time the spatial and the spectral information. $Q_i(\hat{X}^{(k)})$, which is the degree of heterogeneity and can be understood as a degree of membership deficiency, is inversely proportional to the degree of homogeneity and cluster membership. It is defined as the average of the distances of every pixel to the j th class

$$Q_i(\hat{X}^{(k)}) = \frac{1}{L} \sum_{m=1}^L q_j(X_m) \quad (3-5)$$

Where X_m is the m th pixel in the k th neighborhood $\hat{X}^{(k)}$ and L is the number of pixels in that neighborhood. $q_j(X_m)$ is the distance of each pixel to the j th class.

Neighborhood Construction

Overlapped neighborhood structure, the right side shown in Figure 3-12 is the much more effective way to keep and use all the information in an image, therefore it offers the best spatial structure and context information intra or extra a neighborhood, though it is not computation efficient compared with non-overlapped neighborhood structure.

However, the traditional ECHO or UnECHO with non-overlapped neighborhood structure shown

in the left side of Figure 3-12 missed much neighborhood information which was located in the intercross neighborhoods that is shown in the right side of Figure 3-12; subsequently, this is a big impact to their final classification accuracy.

Square neighborhoods were chosen in this implementation due to simplicity and performance considerations. This was not because of restrictions in the theory. The neighborhood sizes are arranged with 3×3 , 4×4 or 5×5 pixels. Neighborhood size has a big impact over the program execution speed. Obviously, larger neighborhoods take more significant time to analyze, at the same time, the amount of neighborhoods in the image is reduced some though. Using the threshold computation method, the amount of thresholds to compute also increases along with the size of the neighborhoods.

Figure 3- 12. 3×6 pixel grid with non-overlapped 3×3 neighborhoods (left),
and with overlapped 3×3 neighborhoods (right).

Estimation of the Threshold TH

STRUC is a vector that represents a specific classification distribution of the pixels in one neighborhood. This classification distribution is a binary image (0 stands for building class, 1 stands for non-building class) based on the results of the supervised segmentation from the first stage on a pixel-by-pixel basis (MSS). For this two-class neighborhood, $\text{STRUC} = [\text{NoBu}, \text{NoNbu}]$, where NoBu is the number of pixels in a neighborhood that after the first-stage classification belongs to the building class, while NoNbu is the number of pixels in a neighborhood that after the first-stage classification belongs to the non-building class. For example, $\text{STRUC} = [2, 7]$ means in a 3×3 neighborhood, 2 pixels belong to building class, while 7 pixels belong to non-building class; but $\text{STRUC} = [7, 2]$ has the reverse meaning.

$\hat{X}^{(k)} \in \text{STRUC}$ means that the particular neighborhood has a STRUC composition in terms of the MSS classification that results from the first stage.

Let $I^{(k)} = \arg \min_i (Q_i(\hat{X}^{(k)}))$. The variable $I^{(k)}$ represents the class to which a neighborhood $\hat{X}^{(k)}$ has the least average distance. For the $\hat{X}^{(k)}$ neighborhood, belonging to $I^{(k)}$ results in that it has the minimum degree of heterogeneity, hence maximum degree of homogeneity.

NS is the number of neighborhoods in the spectral classification result that has the particular

composition STRUC and has the same tendency $I^{(k)}$ belonging to a class. For instance, if in a image, $\hat{X}^{(k)}$ has the STRUC of [2, 7], class tendency $I^{(k)} = 0$; $\hat{X}^{(m)}$ has the STRUC of [2, 7], class tendency $I^{(k)} = 1$; $\hat{X}^{(n)}$ has the STRUC of [3, 6], class tendency $I^{(k)} = 1$; $\hat{X}^{(p)}$ has the STRUC of [2, 7], class tendency $I^{(k)} = 0$; $\hat{X}^{(q)}$ has the STRUC of [2, 7], class tendency $I^{(k)} = 0$, then $NS \{[2, 7], 0\} = 3$, $NS \{[2, 7], 1\} = 1$, $NS \{[3, 6], 0\} = 1$.

QS is the summation of distances to the building class from the neighborhoods that has the same composition STRUC and has the same tendency $I^{(k)}$ belonging to a class.

.

The threshold $TH (STRUC, I^{(k)})$, for the neighborhoods that have the same composition and tendency to a same $I^{(k)}$ class, is computed by

$$TH (STRUC, I^{(k)}) = \alpha \frac{QS(STRUC, I^{(k)})}{NS(STRUC, I^{(k)})} \quad (3-6)$$

where α is threshold coefficient, which is used to modify the threshold value in order to obtain the best segmentation results, whose value varies from 0.9 to 1.3 decided by experiments and the “trial-and-error” rule. QS, NS are the summation of distances and the number of neighborhoods respectively as stated in above. According to Eq.(3-6), the threshold $TH (STRUC, I^{(k)})$ of neighborhoods with the composition structure that has the maximum degree of homogeneity for the I th class is estimated as the average of all these possibilities in the image.

Cross-Neighborhood Structure Detection

However, with the above threshold algorithm, there lies lost information in building roof outlines and a number of misclassified pixels owing to the lack of cross-neighborhood relationship, which is exactly another essential defect in the ECHO or UnECHO. Figure 3-13 illustrates the problem around a trapezoid building (shaded area) in the ECHO or UnECHO.

Figure 3- 13. (a) Various scenarios of neighborhood's locations (type A and type B) around a building (shaded area). (b) The final classification result of the building: many misclassified concave patches and convex patches.

As shown in Figure 3-13, it is obvious that with the ECHO threshold algorithm, many misclassified concave patches and convex patches happen around object boundaries. When the homogeneity tendency $I^{(k)}$ belongs to the non-building class (like 'A' type in the figure), concave patches would probably happen, while the homogeneity tendency $I^{(k)}$ belongs to the

building class (like 'B' type in the figure), convex patches would probably happen. This largely affected the classification accuracy.

Cross-neighborhood Structure Detection (also simplified to STING) takes into account the inter-neighborhood relationship, which is an important improvement to ECHO's intra-neighborhood homogeneity structure analysis.

It has two basic detection tools (STINGs): building STING and non-building STING. The building STING is applied to the targeted pixel whose original class attribute belongs to 'building'. Similarly, the non-building STING is applied to the targeted pixel whose original class attribute belongs to 'non-building'.

First, the 'building' STING is shown in Figure 3-14. For example, when a 4×4 overlapped neighborhood moves to the building boundary, and covers some (minority) building pixels. In this figure, one pixel is targeted for cross-neighbor relationship judgement. Because the total homogeneity in this 4×4 neighborhood is high and its most likely class should be 'non-building', this targeted pixel has to decide whether it goes to 'non-building' class with the neighborhood or stays in 'building' class as the matter of fact without any change. Consequently, as illustrated in the figure, the specific pixel 'tentatively' stretches out four STINGs to detect its cross-neighborhood homogeneity. Once it finds that any one of its four STINGs consists of all

'building' pixels (three), it and its neighborhood will not change anything.

Figure 3- 14. A targeted building pixel stretches out four STINGs to detect its cross-neighborhood homogeneity.

The 'non-building' STING is shown in Figure 3-15. It is relatively complicated with two types: the Main STING and the Secondary STING. Its four STINGs with first-degree connectivity structure (directly contacted connectivity in four straight directions) are called main STINGs (shown in green), whereas its four STINGs with third-degree connectivity (indirectly connectivity in four asymmetrical directions) are called secondary STINGs (shown in light yellow).

If three or four of its four Main STINGs consist of all 'building' pixels (three), the neighborhood

could change all its pixels to its most tendency 'building' class as long as it meets the threshold. If only one or none of the four Main STINGs consists of all 'building' pixels (three), the neighborhood will not change anything even if it meets the change threshold. However, when two of the four Main STINGs consist of all 'building' pixels, the Secondary STINGs are called for further judgement. That is, if two, three or four of the Secondary STINGs consists of all 'building' pixels; the neighborhood could change all its pixels to its most tendency 'building' class as long as it meets the threshold. Otherwise, no changes will be permitted.

Figure 3- 15. A targeted non-building pixel stretches out four main STINGs and four secondary STINGs to detect its cross-neighborhood homogeneity.

Combined with the overlapped neighborhood structure, STINGs greatly improves the final

classification accuracy, and resolves the problem in ECHO and UnECHO – “the overlapped neighborhood would cause iterative operations, result variations and misclassification by convergence” (Jimenez et al., 2005).

3.3.4 Morphological Filtering

Morphological filtering is an effective post-processing algorithm applied to smooth boundaries and to clean unnecessary speckles after the spectral-spatial classification. The ‘closing’ combination of morphological operation is more suitable for building roof extraction as we discussed in Section 3.1.3. It consists of dilation followed by erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries. Thus, the speckle pixels or sporadic outlier pixels in image will be removed, and the major objects become more prominent. Figure 3-16 shows the morphology in 128×128 pixels ‘Townhouse’ image.

(a)

(b)

(c)

Figure 3- 16. Original image (a), buildings before morphological filtering (b), and buildings after morphological filtering (c)

3.3.5 Small Object Filtering

After the morphological filtering, next post-processing is removal of small irrelevant objects in the binary image. Small object footprints, such as individual trees and trucks whose areas are less than $5\text{m} \times 5\text{m}$, are to be deleted. Figure 3-17 shows the effects before and after small object filtering in the same image as Figure 3-16; black arrows show the difference.

Figure 3- 17. Before small object filtering (notice the spots besides arrows) (left); after small object filtering, small impurities disappeared (right).

3.3.6 Building Boundary Extraction and Overlapping

After small object filtering, the last procedure is the boundary extraction using a boundary extraction algorithm (Li, 2004). As a result, in the output image, only building roof outlines are presented, which is obvious and easy to locate and target a specific building for further processing. Finally, in order to show the extracted building in its original environment, the extracted boundaries are overlapped or draped on the original images. Thus, this superposition

gives people a straightforward comparison to tell the extraction accuracy. Figure 3-18 shows the same building roof outlines extracted in Figure 3-17.

Figure 3- 18. Extracted building roof outlines (left), and building roof outlines superposed on original image to visually check the extraction accuracy (right).

3.4 Chapter Summary

In this chapter, an automated spectral-spatial integration extraction approach has been proposed towards 2D building extraction from VHR satellite imagery. First, some theoretical background was given, and then the framework and flowchart of the proposed methodology were described.

With a new Multi-peak Supervised Segmentation (MSS), this approach first generates spectral-based classification. The original colour image and the classification result are then fed into an improved spatial structure classifier. With analysis of the neighborhood's degree of membership to classes, homogeneity and cross-neighborhood structure, the first stage

classification result is revised and improved. Finally, based on the much improved classification result, building roof outlines are extracted. Three unique highlights of this approach are the spectral MSS, spatial Overlapped Neighborhood Analysis, and Cross-neighborhood Structure Detection.

MSS is a new segmentation algorithm and an effective one owing to its easy signature sample selection and high degree of accuracy. The spatial structure classifier was developed from the well known Extraction and Classification for Homogenous Objects (ECHO) classifier. Two new algorithms are exploited in this spatial structure classifier: (1) Overlapped Neighborhood Analysis instead of non-overlapped neighborhood structure in ECHO, and (2) Cross-neighborhood Structure Detection.

In the post-processing, the preceding classification result was improved and enhanced. It includes the ‘closing’ morphological operation and the eliminating of small irrelevant objects, which filter impurities and keep the real important information for the final building extraction. Finally, building roofs are extracted. Moreover, for the comparison with the original building image, the extracted building roof outlines are superposed on the original image for visual evaluation.

Chapter 4

IMPLEMENTATION

This chapter describes the implementation undertaken on the new sequential spectral-spatial classification approach. The comparison between the proposed algorithms and other algorithms are reported according to the experimental results, including unsupervised ISODATA segmentation, supervised SE classification, ECHO or UnECHO spectral-spatial sequential classification are chosen for analysis and comparison. Also, different neighborhood sizes and different homogeneity threshold coefficients are tested.

4.1 Datasets and Software Environment

The implementation objects are typical urban buildings, like townhouses, a community centre, storage buildings, industrial buildings and apartments. These 1m resolution IKONOS satellite images were downloaded from a public website (<http://earth.google.com/>), and the main source image was named 'DLR-gelaende_8706×4860_1m.JPG'. Most images for implementation are cropped from it. The following images are some typical and representative images among many images implemented in this study. They will satisfactorily demonstrate the performance of the proposed approach:

(1) 128×128 pixels images: Figure 4-1 (a), 1m resolution, townhouses, the neighborhood size and threshold coefficient (threshold coefficient is defined in Section 3.3.3, equation 3-6) used in this implementation are 3×3 and 1.15, respectively. Figure 4-1 (b), 1m resolution, a community centre, the neighborhood size and threshold coefficient used in implementation are 4×4 and 1.20, respectively.

(2) 150×150 pixels images: Figure 4-1 (c), 1m resolution, industrial buildings, the neighborhood size and threshold coefficient used in this implementation are 3×3 and 1.10, respectively. Figure 4-1 (d), 1m resolution, apartments, the neighborhood size and threshold coefficient used in implementation are 3×3 and 1.15, respectively.

Figure 4- 1. Four test images used in the study (a) townhouses, (b) community centre, (c) industrial and (d) apartment buildings.

In this study and experiments, MatLAB is used as a main programming analysis tool. One of its import platforms is for image processing and analysis. Users can extract, analyze and compare

any imagery data based on C++ and other languages. Moreover, it has strong capability for complicated mathematical calculation, friendly programming interface and convenient accesses to tabular data via interactive query. Dozens of programs for this proposed approach are based on MatLAB and some major ones are presented in the appendix at the end of this thesis.

Nevertheless, in order to make objective comparison with the proposed new approach, commercial software PCI Geomatica V9.1-FOCUS and ERDAS Imagine are also used in ISODATA segmentation and traditional supervised SE classification (i.e., Signature Editor in ERDAS). A 1.50 GHz Processor, 512MB of RAM desktop computer is the major hardware for this implementation calculation and program operation.

4.2 Experiments

4.2.1 Spectral Segmentation Comparison between MSS and Unsupervised ISODATA

Spectral Segmentation

ISODATA is one of the well known sophisticated unsupervised segmentation techniques. However, the experimental results illustrated its drawbacks. In Figure 4-2 (buildings are all in red), column (a) are original images, column (b) are the results from ISODATA, and column (c) are the results from MSS. By visual inspection, we can quickly tell that the better ones are from

the proposed MSS approach, especially in those locations denoted by black arrows. It is easy to tell the apparent differences. Further more detailed qualitative and quantitative statistics assessment will be presented in Chapter 5.

Apartments

Figure 4- 2. Spectral segmentation comparison. Original images (a), Results of ISODATA (b), and Results of MSS (c).

In the “Townhouses” image, ISODATA had poor building segmentation due to the “pepper and salt” effect and a large area of road was also mistakenly extracted as building (denoted by the black arrow). In the “Industrial buildings” image, buildings were extracted in ISODATA, but compared with MSS, many buildings had faulty shape and lost area, especially the adjoining buildings were almost separated (denoted by the black arrow). In the “Apartments” image, ISODATA had the similar problem indicated by the black arrow.

Spectral-spatial integration result

In order to see more distinguishing differences between MSS and ISODATA, the above spectral segmentation results were further fed into spatial information analysis by going through the same proposed spectral-spatial integration classification approach. The integration approach results are in binary images: buildings are black, thus it is easier for people to compare the above

spectral-based classification approaches.

In Figure 4-3, column (a) are original images, column (b) are the spectral-spatial integration classification based on ISODATA spectral result, and column (c) are the spectral-spatial integration classification based on MSS spectral result. (Notice the similar red arrows)

Apartments

Figure 4- 3. Further comparison between ISODATA and MSS. Original images (a);

Spectral-spatial integration classification based on ISODATA spectral result (b); Spectral-spatial integration classification based on MSS spectral result (c).

4.2.2 Spectral segmentation comparison between MSS and Signature Editor (SE)

Spectral Segmentation

Considering that possible doubts may be raised because the previous comparison is between a supervised segmentation (MSS) and unsupervised segmentation (ISODATA), which seems not fair. As a result, in this section, to guarantee an unbiased comparison, traditional supervised segmentation approach SE is also used in comparison with MSS. As aforementioned, SE is one of the typical sophisticated supervised segmentation techniques.

In Figure 4-4, column (a) are original images, column (b) are the SE results, and column (c) are

the MSS results. Buildings are all in red.

Apartments

Figure 4- 4. Comparison between SE and MSS. Original images (a); SE results (b);
MSS results (c).

By visual inspection, we can tell the better results are from the proposed MSS approach, especially those locations denoted by black arrows. In the “Townhouses” and “Community center” images, SE had difficulties in following roof outline extraction due to its loss of building information; In the “Industrial buildings” image, buildings were barely extracted in SE, compared with MSS, many buildings have faulty shape and lost area, especially the adjoining buildings which were almost separated (denoted by the black arrow). In the “Apartments” image, SE had the problem of missing buildings indicated by the black arrow.

Spectral-spatial Integration Result

Similarly, in order to see more distinguishing differences between MSS and SE, the above spectral segmentation results were further fed into spatial information analysis by going through the same proposed spectral-spatial integration classification approach. The integration approach

results are in binary images with buildings are black, thus it is easier for people to compare the above spectral-based MSS and SE.

In Figure 4-5, Notice those red arrows showing the problems of lost information or missing buildings in SE. Column (a) are original images, column (b) are the spectral-spatial integration classification based on SE spectral result, and column (c) are the spectral-spatial integration classification based on MSS spectral result.

Industrial
buildings

Apartments

(a)

(b)

(c)

Figure 4- 5. Further comparison between SE and MSS. Original images (a); Spectral-spatial integration classification based on SE spectral result (b); Spectral-spatial integration classification based on MSS spectral result (c).

4.2.3 Comparison between Spectral-only and Spectral-spatial Classification

In the previous sections, the superiority of MSS was illustrated by comparison with both unsupervised and supervised traditional segmentation methods. In this section, it is essential to explain the advantages of integrating spectral and spatial information by the proposed approach. First of all, the spectral-only classification results will be presented, then, the proposed

spectral-spatial integration classification will be given as the comparison counterpart. From the following extraction results (Figure 4-6), it is not difficult to tell the proposed spectral-spatial integration classification has obvious advantages over a spectral-only classification. In Figure 4-6, to enhance the visual comparison accuracy, ground truth (red outlines) also were introduced in each result, whereas black outlines represent extracted building roof outlines, and (a) are original images; (b) are spectral-only building extractions; (c) are spectral-spatial building extractions; (d) are spectral-only extractions and ground truth; and (e) are spectral-spatial extractions and ground truth.

(a)

(b)

(c)

Community centre

(d)

(e)

(a)

(b)

(c)

Industrial buildings

Industrial buildings

(d)

(e)

(a)

(b)

(c)

Apartments

(d)

(e)

Figure 4- 6. Comparison between spectral-only and spectral-spatial integration. Original images (a); Spectral-only extractions (b); Spectral-spatial extractions (c); Spectral-only extractions and ground truth (d); Spectral-spatial extractions and ground truth (e).

Illustrated by those locations near the black arrows, the experiment results showed that the integration of spectral-spatial information is more accurate and meaningful to the human analysis than pixel-by-pixel spectral-only segmentation.

4.2.4 Comparison between proposed spectral-spatial approach and ECHO

ECHO is a well known sophisticated spectral-spatial approach. However, it has some weakness aforementioned. In this proposed spectral-spatial approach, three new methods are used to improve its accuracy in building extraction. In ECHO or UnECHO, there is no MSS in its spectral analysis stage, and there are no Overlapped Neighborhood Analysis (ONA) and Cross-neighborhood Structure Detection (CSD) in its second stage. Consequently, the obvious differences can be visually distinguished in extraction results as follows (red outlines represent ground truth, black outlines represent final building roof outlines). Further more detailed qualitative and quantitative statistics analysis and evaluation will be presented in Chapter 5.

In Figure 4-7, ground truth (red outlines) were also introduced in each result, and (a) are original images; (b) are building extractions by ECHO; (c) are building extractions by proposed approach; (d) are ECHO extractions and ground truth; and (e) are proposed approach extractions and ground truth.

(a)

(b)

(c)

Townhouses

(d)

(e)

(a)

(b)

(c)

Community centre

(d)

(e)

(a)

(b)

(c)

Industrial buildings

(d)

(e)

(a)

(b)

(c)

Apartments

(d)

(e)

Figure 4- 7. Comparison between ECHO and proposed integration approach. Original images (a); ECHO building extractions (b); Proposed approach extractions (c); ECHO extractions and ground truth (d); Proposed approach extractions & ground truth (e).

By visual inspection, we can tell the better results are from the proposed approach. They have more precise building area and boundary extractions, which match the ground truth better than ECHO. Especially, those locations denoted by black arrows definitely point out the differences between the two approaches. In the “Townhouses” and “A community center” images, proposed approach had more accurate and plump roof outline extractions with less building information loss. In the “Industrial buildings” images, buildings were well extracted in both approaches, but with the ECHO result, the adjoining buildings were almost separated (denoted by the black arrow). In the “Apartments” images, ECHO had similar defects.

4.2.5 Comparison between Different Neighborhood Sizes

Overlapped neighborhood structure keeps and uses more image information, but it is more computation intensive compared with non-overlapped neighborhood structure. Neighborhood size has a big impact over the program execution speed. Larger neighborhood size takes more significant time to analyze. Not only because there are more pixels to be analysis in bigger neighborhood size, but also because the neighborhood structure type analysis increases as the

neighborhood size increases.

Furthermore, neighborhood size has some influence on the accuracy of the final extraction result. To examine this, the following (Figure 4-8 and Figure 4-10) are the neighborhood size experimental tests arranged with 3×3 , 4×4 or 5×5 , 6×6 various neighborhood sizes, with a 1.50 GHz Processor, 512MB of RAM desktop computer. In each figure, column (a) shows the final extraction results, column (b) shows the final results superposed with original images.

(1) Figure 4-8. The community centre (128 × 128 pixels):

Neighborhood size: 5×5

Neighborhood size: 6×6

(a)

(b)

Figure 4- 8. Different accuracy with different neighborhood sizes in “Community centre”. Final extraction results in Column (a), and final results superposed on original images in column (b).

However, execution time (55 minutes, 90 minutes, 135 minutes, and 190 minutes) with respective neighborhood size (3×3 , 4×4 , 5×5 , and 6×6) is shown in Figure 4-9.

Figure 4- 9. Different execution time with different neighborhood sizes.

From Figure 4-8, we can see the extraction accuracy did not increase with the increase of neighborhood size. By the comparison of these results (particularly the spots denoted by black arrows), the best neighborhood size for this “community center” image is 4×4 with regard to accuracy. Moreover, from Figure 4-9, it is found out that the execution time will increase exponentially with the increase of size. Therefore, regarding both the accuracy and execution time, size 4×4 is the premium neighborhood size.

(2) Figure 4-10. Storage buildings (150 × 150 pixels):

Neighborhood size: 5×5

Neighborhood size: 6×6

(a)

(b)

Figure 4- 10. Different accuracy with different neighborhood sizes in “Storage buildings”. Final extraction results in Column (a), Final results superposed with original images in column (b).

Similarly, execution time (75 minutes, 133 minutes, 208 minutes, and 300 minutes) with respective neighborhood size (3×3 , 4×4 , 5×5 , and 6×6) is shown in Figure 4-11.

Figure 4- 11.
Different execution
time with different
neighborhood sizes.

From Figure 4-10, we can see the extraction accuracy did not increase with the increase of neighborhood size. By the comparison of these results (particularly the spots denoted by black arrows), the best neighborhood size for this “storage buildings” image is 3×3 with regard to accuracy. Moreover, from Figure 4-11, it is found out that the execution time increases exponentially with the increase of size. Therefore, regarding the accuracy and execution time, size 3×3 is the premium neighborhood size.

In summary, with respect to the above comparison and illustration, together with many other different size image tests, neighborhood size 3×3 and 4×4 are proved to be the optimal size both for the accuracy and program execution speed.

4.2.6 Comparison between Different Thresholds

As stated in Section 3.3.3, the homogeneity threshold is the key for judging whether a neighborhood needs to change its class as a whole or not according to its homogeneity degree of membership. Therefore, the homogeneity threshold has influence on the final accuracy of extraction. Threshold coefficient α (see Section 3.3.3, Eq. 3-6) is an essential factor in the threshold. The following (Figure 4-12 and Figure 4-13) are the threshold coefficient α tests by modifying its different value ranging from 0.9 to 1.3 in an error-and-trial way. In each figure, column (a) shows the final extraction results, column (b) shows the final results superposed with original images.

(1) Figure 4-12. The community centre (128 × 128 pixels):

$\alpha = 1.15$

$\alpha = 1.2$

$\alpha = 1.3$

Figure 4- 12. Threshold coefficient comparison in “Community center”.

(2) Figure 4-13. The storage buildings (150 × 150 pixels):

(a)

(b)

$\alpha = 0.9$

$\alpha = 1.0$

$\alpha = 1.05$

$\alpha = 1.1$

$\alpha = 1.15$

$\alpha = 1.2$

$\alpha = 1.3$

Figure 4- 13. Threshold coefficient comparison in “Storage buildings”.

In Figure 12 and Figure 13, following the black arrows, we can see the gradual change as the threshold coefficient α changes. $\alpha = 1.2$ is the best one for “community centre” image extraction, whereas $\alpha = 1.05$ is the best one for “storage buildings” image extraction. By comparing Figure 12 and Figure 13 with Figure 8 and Figure 10, it can be seen that threshold coefficient has less influence than neighborhood size. This testing shows that 1.05-1.20 is the optimal threshold coefficient range.

4.3 Chapter Summary

By visual comparison and analysis, it can be seen that significant qualitative differences exist between the proposed algorithms and other algorithms, including unsupervised ISODATA segmentation, supervised SE, ECHO or UnECHO spectral-spatial sequential classification approach. The proposed approach has obvious advantages over them, both in spectral stage and spectral-spatial stage classification. Also the buildings extracted in these images are very pleasing. Through experiments, it was also found that neighborhood size 3×3 or 4×4 are the optimal size both for accuracy and program execution speed and 1.05-1.20 is the optimal threshold coefficient range.

Chapter 5

RESULTS AND QUALITY EVALUATION

This chapter introduces some qualitative and quantitative evaluation methodologies for the accuracy assessment of experiment results. By using these evaluation methodologies, the effectiveness and superiority of the proposed approach are further investigated by comparison with ECHO and UnECHO, ISODATA, SE.

5.1 Generation of Ground Truth Images

Accuracy assessment is a general term for comparing the classification to geographical data that are assumed to be true. Usually, the assumed-truth data are derived from ground truth data.

Ground truth data is sometimes the baseline images showing the real status, location and shape of those image objects. They are the basis for qualitative and quantitative evaluation of image extraction. Usually there are two ways to generate them, the first evolves making good use of existing or previous geo-data to generate them with field surveying; the second is manual measurement in the colour orthoimages in GIS software. As for the satellite images downloaded from the Internet, usually the second method is chosen.

In this study, the ground truth data was produced by manual measurement and extraction in the colour orthoimages in ArcGIS 9.0 software, by an experienced operator. Figure 5-1 shows sample images and the ground truth that was manually measured from the images.

Figure 5- 1. Ground truth examples. (a) & (b) Townhouses and its ground truth, (c) & (d)

Industrial buildings and its ground truth

5.2 Evaluation Methodology

The basic idea is to compare the predicted classification (supervised or unsupervised) of each pixel with the actual classification as discovered by ground truth. The misidentification of individual features (thematic error) is our detecting object. The evaluation methodology usually includes visual comparison, the confusion matrix and Kappa statistics (κ), and the building extraction rate.

In Chapter 4, the effectiveness of the proposed approach has been illustrated by visual evaluation.

In this chapter, focus will be concentrated on qualitative and quantitative accuracy statistics, i.e., the confusion matrix, κ statistics, and the building extraction rate.

Confusion Matrix and Kappa Statistics

The most common means of expressing classification accuracy is the confusion matrix, which provides a comparison on a category-by-category basis of classification results compared with the known reference data (ground truth). Table 5-1 is an example showing Confusion Matrix accuracies and Kappa statistics (κ). Those numbers in the table are the correctly extracted object pixels and mistakenly extracted object pixels.

Table 5- 1. A confusion matrix example (unit: pixel)

Classification	Water (W)	Concrete (C)	Buildings (B)	Row total
Water (W)	93	0	4	97
Concrete (C)	0	65	5	70
Buildings (B)	2	3	165	170
Column total	95	68	174	337

From the above table, some evaluation index can be calculated as follows,

Producer's accuracy: $W = 93/95 = 98\%$, $C = 65/68 = 96\%$, $B = 165/174 = 95\%$

User's accuracy: $W = 93/97 = 96\%$, $C = 65/70 = 93\%$, $B = 165/170 = 97\%$

Overall accuracy: $(93 + 65 + 165)/337 = 95.8\%$

$\kappa = (323 - 129.53) / (337 - 129.53) = 93.3\%$

The following will explain the above calculations:

In a confusion matrix (see Table 5-1), the major diagonal numbers are proper assessed pixels, while the non-diagonal elements are errors. There are two types of errors in the confusion matrix: first, the Omission (exclusion, in down columns) means the pixels, which are supposed to be included, but which were indeed excluded. In the above example, like column 'water' omission error is $2/95$, i.e., 2.1%. Its opposite is the Producer's Accuracy, 97.9%. The second error type is Commission (erroneous inclusion), along row direction, which means including a pixel in a class when it should have been excluded, like row 'concrete' commission is $5/70$, i.e., 7%. Its opposite is the User's Accuracy, 93%.

Overall Accuracy in a confusion matrix means the accuracy dividing the total number of correctly classified pixels (the major diagonal numbers) by the total number of reference pixels.

In the above example, it is $(93 + 65 + 165)/337 = 95.8\%$.

κ statistics expresses the proportionate reduction in error generated by a classification process, compared with the error of a completely random classification. For perfect agreement, $\kappa = 1$,

while a value of 0.82 would imply that the classification process was avoiding 82 % of the errors that a completely random classification would generate.

$$\kappa = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \cdot x_{+i})} \quad (5-1)$$

where, r: # of rows in error matrix

N: total # of observations in error matrix

x_{ii} : major diagonal element for class I

x_{i+} : total # of observations in row i (right margin)

x_{+i} : total # of observations in column i (bottom margin)

Building Extraction Rate

The building extraction rate used in this work has been described and employed elsewhere (Shufelt, 1996; Fischler et al., 1998; Li, 2000), although the specific definitions used in previous work have varied. The definition used in this work is described in detail here to avoid confusion; it measures the matching degree of the extracted building to its ground truth. The building extraction rate is

$$\text{Building extraction rate} = \frac{BT}{BU + TR + BT} * 100\% \quad (5-2)$$

where, BT is the number of pixels locating within both the extracted buildings and the ground

truth boundaries; and BU is the number of pixels falling within the extracted buildings' boundaries only; TR is the number of pixels falling within the ground truth's boundaries only.

5.3 Evaluation of Results and Discussion

As for the random reference pixels choice in the confusion matrix, the number of reference pixels is an important factor in determining the accuracy of the classification. It has been shown that more than 250 reference pixels are needed to estimate the mean accuracy of a class to within plus or minus 5% (ERDAS, 1991). In this study, the confusion matrix is based on 250 randomly chosen pixels produced by the random number generator in Microsoft EXCEL.

5.3.1 Confusion Matrix and Kappa Statistics

The previous experiment results will be evaluated by the confusion matrix and κ statistics in three kinds of comparisons: A. spectral classification, B. spatial-spectral versus spectral-only classification, and C. proposed approach versus ECHO.

A. The spectral-based classification comparison among MSS, unsupervised ISODATA and supervised SE method.

In order to see the differences among them more clearly, the spectral classification results were further fed into the proposed spectral-spatial integration approach, so outputs are final

roof outlines (in black). The following (Figure 5-2 and Figure 5-3) show the final results by using ISODATA, SE, and MSS. Column (a) show results compared with ground truth (in red); column (b) show results (in blue) superposed on original image.

(1) Figure 5-2. The ‘Townhouses’ image:

Figure 5- 2. “Townhouses” spectral-based classification comparison among MSS, ISODATA and SE. Results with ground truth (a), and results superposed on original image (b).

Table 5- 2. Confusion matrix of ISODATA, SE, MSS results

ISODATA results				
Classification	Building	Non-building	row total	User's accuracy
Building	31	32	63	49.2%
Non-building	21	166	187	
Column total	52	198	250	
Producer's accuracy	59.6%			
Overall accuracy	78.8%			
κ	40.3%			
SE results				
Classification	Building	Non-building	row total	User's accuracy
Building	30	2	32	93.8%
Non-building	29	189	218	
Column total	59	191	250	
Producer's accuracy	50.8%			
Overall accuracy	87.6%			
κ	59.2%			
MSS results				
Classification	Building	Non-building	row total	User's accuracy
Building	39	5	44	88.6%
Non-building	7	199	206	
Column total	46	204	250	
Producer's accuracy	84.8%			
Overall accuracy	95.2%			
κ	83.7%			

(2) Figure 5-3. The ‘Industrial buildings’ image:

Figure 5- 3. “Industrial buildings” spectral-based classification comparison among MSS, ISODATA and SE. Results with ground truth (a), and results superposed on original image (b).

Table 5- 3. Confusion matrix of ISODATA, SE, MSS results

ISODATA results				
Classification	Building	Non-building	row total	User's accuracy
Building	48	2	50	96.0%
Non-building	17	183	200	
Column total	65	185	250	
Producer's accuracy	73.8%			
Overall accuracy	92.4%			
κ	78.7%			
SE results				
Classification	Building	Non-building	row total	User's accuracy
Building	47	2	49	95.9%
Non-building	13	188	201	
Column total	60	190	250	
Producer's accuracy	78.3%			
Overall accuracy	94.0%			
κ	82.5%			
MSS results				
Classification	Building	Non-building	row total	User's accuracy
Building	62	7	69	89.9%
Non-building	2	179	181	
Column total	64	186	250	
Producer's accuracy	96.9%			
Overall accuracy	96.4%			
κ	90.8%			

From the above confusion matrix and Kappa Statistics (κ) comparison, it can be seen that MSS significantly improved the κ , Producer's accuracy and overall accuracy. Especially, the essential

κ was improved noticeably: more than 50% accuracy increase (from 40% to 83%) in “Townhouses” image; 15% accuracy increase (from 78.7% to 90.8%) in “industrial buildings”.

B. Comparison between spectral-only and spectral-spatial

Figure 5- 4. Spectral-only and spectral-spatial results. (a). Results with ground truth (in red) (b); and results (in blue) superposed on original image (c).

Table 5- 4. The confusion matrix of spectral-only and spectral-spatial segmentation

Spectral-only				
Classification	Building	Non-building	row total	User's accuracy
Building	54	6	60	90.0%
Non-building	10	180	190	
Column total	64	186	250	
Producer's accuracy	84.4%			
Overall accuracy	93.6%			
κ	82.8%			
Spectral-spatial				
Classification	Building	Non-building	row total	User's accuracy
Building	62	7	69	89.9%
Non-building	2	179	181	
Column total	64	186	250	
Producer's accuracy	96.9%			
Overall accuracy	96.4%			
κ	90.8%			

From the above confusion matrix and Kappa Statistics (κ) comparison, it can be seen that by integrating spectral and spatial information instead of only using spectral information, the accuracy was apparently increased, especially, the κ increased 9.7% (from 82.8% to 90.8%).

C. Comparison between ECHO and proposed spectral-spatial approach

In order to clearly illustrate the differences among them, firstly, two new algorithms in this proposed approach were picked out to compare with ECHO individually. Thus, the benefits of these two new algorithms in the proposed approach can be seen more distinctly. Figure 5-5 shows the spectral-spatial integration results using ECHO, ECHO+ ONA, ECHO+ CSD, and the proposed approach. Column (a) show the results compared with ground truth (in red), and column (b) show the results (in blue) superposed on the original images.

Proposed approach

Figure 5- 5. ECHO, ONA, CSD and the proposed approach comparison. Results with ground truth (in red) (a), and results (in blue) superposed on original image (b).

Table 5- 5. The confusion matrix of ECHO, ECHO+ ONA, ECHO+ CSD, proposed approach classification results

ECHO				
Classification	Building	Non-building	row total	User's accuracy
Building	38	4	42	90.5%
Non-building	18	191	209	
Column total	56	195	251	
Producer's accuracy	67.9%			
Overall accuracy	91.2%			
κ	72.2%			
ECHO+ ONA				
Classification	Building	Non-building	row total	User's accuracy
Building	39	4	43	90.7%
Non-building	14	193	207	
Column total	53	197	250	
Producer's accuracy	73.6%			
Overall accuracy	92.8%			
κ	76.9%			

ECHO+ CSD				
Classification	Building	Non-building	row total	User's accuracy
Building	40	5	45	88.9%
Non-building	14	191	205	
Column total	54	196	250	
Producer's accuracy	74.1%			
Overall accuracy	92.4%			
κ	76.1%			
Proposed approach				
Classification	Building	Non-building	row total	User's accuracy
Building	39	5	44	88.6%
Non-building	7	199	206	
Column total	46	204	250	
Producer's accuracy	84.8%			
Overall accuracy	95.2%			
κ	83.7%			

From the above comparison, it can be seen that Overlapped Neighborhood Analysis (ONA) and Cross-neighborhood Structure Detection (CSD) contribute about 4%-5% individually to the final improved accuracy (83.7%), compared with ECHO (72.2%).

D. Summary of confusion matrix evaluations:

Table 5- 6. Summary of Overall Accuracy and the κ evaluation

Building type	Accuracy index type	ISODATA	SE	ECHO	ECHO + ONA	ECHO + CSD	PROP. APPR.
Town-house	Overall	78.8%	87.6%	91.2%	92.8%	92.4%	95.2%
	κ	40.3%	59.2%	72.2%	76.9%	76.1%	83.7%
Industrial-buildings	Overall	92.4%	94.0%	94.3%	95.1%	94.8%	96.4%
	κ	78.7%	82.5%	84.2%	85.6%	85.5%	90.8%

From the above results, according to κ and an overall accuracy evaluation, the following conclusions can be drawn:

- 1) In spectral segmentation and classification, supervised MSS (83%) is more effective than unsupervised ISODATA (40%). In particular, it is also superior to traditional supervised classification (SE) (59%).
- 2) In the spectral-spatial integration, based on MSS results, both new algorithms- Overlapped Neighborhood Analysis and Cross-neighborhood Structure Detection contribute to the significantly improved accuracy (83%), compared with the well known ECHO (72%). The proposed approach is tested effective with ONA, CSD and MSS.

- 3) Tedious traditional supervised classification (SE) is not as effective as people expected, which explains that sometimes people prefer unsupervised segmentation to traditional supervised classification.

5.3.2 Building Extraction Rate

A. Spectral comparison among MSS, unsupervised ISODATA and supervised SE:

Figure 5-6 shows the spectral results from ISODATA, SE, and MSS. Column (a) shows results compared with ground truth (in red); column (b) show results (in blue) superposed on the original image.

MSS

(a)

(b)

Figure 5- 6. Spectral comparison among MSS, ISODATA and SE. Results compared with ground truth (a); results superposed on original image (b).

Table 5- 7. The extraction rate of ISODATA, SE, and MSS

Type	BU	TR	BT
ISODATA			
Pixels	32	21	31
Extraction Rate	36.9%		
SE			
Pixels	2	29	30
Extraction Rate	49.2%		
MSS			
Pixels	5	7	39
Extraction Rate	76.5%		

From the above extraction rates, it can be seen that MSS had the outstandingly high extraction rate of 76.5%, compared with ISODATA at 36.9% and SE at 49.2%.

B. Comparison between ECHO and proposed spectral-spatial approach:

Figure 5-7 and Figure 5-8 show the spectral-spatial integration results from ECHO and the proposed approach. Column (a) shows the results compared with ground truth (in red); column (b) shows the results (in blue) superposed on the original image.

(1) Figure 5-7. The ‘Townhouses’ image:

Figure 5- 7. Comparison between ECHO and proposed approach in “Townhouses”. Results compared with ground truth (a); Results (in blue) superposed on original image (b).

Table 5- 8. The extraction rate of ECHO and proposed approach

Type	BU	TR	BT
ECHO			
Pixels	4	18	38
Extraction Rate	63.3%		
Proposed approach			
Pixels	5	7	39
Extraction Rate	76.5%		

(2) Figure 5-8. The ‘Industrial buildings’ image:

Figure 5- 8. Comparison between ECHO and proposed approach in “Industrial buildings”.

Results compared with ground truth (a); Results superposed on original image (b).

Table 5- 9. The extraction rate of ECHO and proposed approach

Type	BU	TR	BT
ECHO			
Pixels	5	8	56
Extraction Rate	81.2%		
Proposed approach			
Pixels	7	2	62
Extraction Rate	87.3%		

C. Summary of Extraction Rate evaluation:

Table 5- 10. The summary of extraction rate comparison

Type \ Methods	Extraction Rate			
	ISODATA	SE	ECHO	PROPOSED
Townhouse	36.9%	49.2%	63.3%	76.5%
Industrial Buildings	57.4%	69.8%	81.2%	87.3%

From the above extraction rate comparison, it can be seen that the proposed approach improves the building extraction accuracy, increasing “Townhouse” from 36.9% (ISODATA) to 63.3% (ECHO), and to itself 76.5%, whereas “Industrial buildings” from 57.4% (ISODATA) to 81.2% (ECHO), and to itself 87.3%. Experiments strongly indicate the effectiveness of the three new algorithms-MSS, ONA and CSD in this proposed approach.

5.3.3 Factors Affecting the Accuracy of Results

(1) Effect of low image contrast: image contrast has a great effect on building extraction accuracy. By testing many images with various contrast, the author found that if the contrast of the image is low, the extraction process might fail to extract building outlines precisely. The effect is obvious in VHR satellite images. Usually the low contrast may cause the building to be under or over extracted and thus affect the geometric accuracy of the extracted building.

(2) Effect of image noise: the radiometric quality of VHR satellite images has an effect on the final building extraction results. Although noise minimization methods can be applied to minimize this effect, it is necessary to be aware the noise impact to the building extraction which leads to loss of small, but still important, building details. In general, image noise can affect interpretation and then you get extraction with incorrect building outlines.

(3) Effect of reflectance and shadow: various sun angles and shadowing have disturbing effects in areas where buildings are very close to each other and building roofs have similar colour with shadows. The influence of reflectance was obvious on IKONOS images where shadows were confused as building candidates and falsely extracted.

Chapter 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Contributions

In this thesis, an automated spectral-spatial integration extraction approach has been developed for extracting 2D urban buildings from VHR IKONOS imagery. With three new algorithms-MSS, ONA, and CSD, this approach has a significant improvement in Extraction Rate and Kappa Index accuracy, compared with the well known ECHO, ISODATA and SE.

With the spectral-based Multi-peak Supervised Segmentation (MSS), the approach generates first stage classification. In its second stage, the original colour image and the first stage results are then fed into a spatial structure classifier. With analysis of neighborhood's degree of membership to classes, spatial homogeneity and cross-neighborhood structure, the first stage classification is significantly improved. Finally, buildings are extracted with a post-processing. Compared with ISODATA and ECHO, three contributions are the spectral MSS, spatial Overlapped Neighborhood Analysis (ONA) and spatial Cross-neighborhood Structure Detection (CSD).

The innovative spectral MSS gives more accurate spectral classification in the first stage. It was proven to be a sound platform for further spectral-spatial information integration. According to the classification results, supervised MSS ($\kappa = 83\%$) is much more effective than unsupervised ISODATA ($\kappa = 40\%$). In particular, it is also much superior to traditional supervised classification SE ($\kappa = 59\%$). With the precise feature signature capture technology in MSS, it significantly simplified the tedious signature selection work that has been such a problem in SE, and meanwhile the accuracy was increased substantially.

Because the ONA is much more spatially informative than the non-overlapped neighborhood structure in ECHO, and CSD has unique STING with strong error-correcting abilities, an improved spatial structure classifier with the ONA and CSD presented much more effective and meaningful classification results by integrating the spectral and spatial information. In the second stage, i.e., spectral-spatial integration, both the innovated ONA and CSD contributed to the well-improved accuracy ($\kappa = 83\%$), compared with the well-known ECHO ($\kappa = 72\%$). Similarly, it also improved the building extraction accuracy substantially (increased from 63.3% to 76.5% and 'Industrial buildings' from 81.2% to 87.3%), strongly demonstrating the effectiveness of MSS, ONA and CSD.

In summary, this study successfully developed an effective spectral-spatial sequential integration approach to extract 2D urban buildings from VHR IKONOS imagery. More specifically, the

proposed approach has noticeably higher extraction accuracy compared with ECHO, ISODATA, and ES. It is a success; consequently, the object of this study was fulfilled.

6.2 Recommendations for Future Research

For building extraction, although the proposed spectral-spatial approach demonstrates promising results, several priorities for future research remain.

First, although the neighborhood homogeneity spatial structure was fully used in this approach, but the texture information was not employed. Presently, there are alternative methods for describing colour texture (Paschos and Petrou, 2003; Mäenpää and Pietikäinen, 2004; Palm, 2004). Integrating the texture information will further improve the results, thus, further experiments on using texture descriptor and this spatial structure classifier will help determine better feature measures for segmentation.

Second, this approach was only applied on common spectral images, but has not been tested on hyperspectral data. Good results might be achieved as well, but how to fuse different band's values and consider cross-band relations is a new problem for future research. Further analyses and experiments are necessary for determining an optimal balance for combining them.

Third, the general future research direction should be the fusion of multi-source satellite imagery to extract more accurate and detailed building images, as well as digital elevation/surface information (e.g., airborne Lidar or InSAR) for a more intelligent and robust perception. Integrating the detection results from different image levels, including contextual information, can also provide additional evidence for forming a more complete building outline, and help improve the results. The mechanism need to combine and integrate all those information is one of the key problems and challenges for the future work.

REFERENCES

- Asano, T., D.Z. Chen, N. Katoh, and T. Tokuyama, 1996. Polynomial-time solutions to image segmentation. www.portal.acm.org, March, 2003.
- Batistella, M., S. Robeson, and E.F. Moran, 2003. Settlement Design, Forest Fragmentation, and Landscape Change in Roddonia, Amazonia. *Photogrammetric Engineering & Remote Sensing*, 67(5): 571-574.
- Benz, U.C., P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen, 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry & Remote Sensing*, 58: 239-258.
- Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems, *Journal of Royal Statistics Society B*, 36: 192-236.
- Besag, J., 1986. On the statistical analysis of dirty pictures. *Journal of Royal Statistics Society B*, 48: 369-434.
- Broomhead, D. S. and D. Lowe, 1988. Multivariable function interpolation and adaptive networks, *Complex Systems*, 2: 321-355.
- Caelli, T., and D. Reye, 1996. On the classification of image regions by colour, texture and shape, *Pattern Recognition*, 26:461-470.
- Canny, J., 1986. A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8, 6:679-698.

- Carleer, A.P., O. Debeir, and E.Wolff, 2005. Assessment of Very High Spatial Resolution Satellite Image Segmentations. *Photogrammetric Engineering & Remote Sensing*, 71, (11): 1285-1294.
- Chellappa, R., S. Chatterjee and R. Bagdazian, 1985. Texture synthesis and compression using Gaussian-Markov random field Models, *IEEE Transactions on Systems Man and Cybernetics*, 15 (2): 298-303.
- Chellappa, R. and A. K. Jain, 1993. Markov Random Fields, Theory and Applications, Academic Press, San Diego.
- Chen, K.M., and S.Y. Chen, 2002. Colour texture segmentation using feature distributions. *Pattern Recognition Letters*, 23: 755–771.
- Cheng, H.D., X.H. Jiang, Y. Sun, and J. Wang, 2001. Colour image segmentation: advances and prospects, *Pattern Recognition*, 34:259-281.
- Cohen, F., 1986. Markov random fields for image modelling and analysis, Desai, U. (Ed.) *Modelling and Application of Stochastic Processes*, 243-272, Kluwer Academic, Boston.
- Cohen, F. and D. Cooper, 1987. Simple parallel hierarchical and relaxation algorithms for segmenting noncausal Markovian random field, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9 (2): 195-219.
- Cushnie, J.L., 1987. The interactive effects of spatial resolution and degree of internal variability within land-cover type on classification accuracies. *International Journal of Remote Sensing*, 8: 152-161.

- Definiens, 2003. eCognition Professional, Version 3.0, Munich: Definiens Imaging Gmbh, www.definiedns-imaging.com.
- Delia, C., G. Poggi, and G. Scarpa, 2003. A tree-structured Markov random field model for Bayesian image segmentation. *IEEE Transactions on Image Processing*, 12 (10): 1259 - 1273.
- Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistics Society B*, 39: 1-8.
- DeValois, R. and K. DeValois, 1990. *Spatial Vision*, Oxford University Press, Oxford
- DeYoe, E. and D. van Essen, 1996. Concurrent processing streams in monkey visual cortex, trends, *Neurosciences*, 11: 219-226.
- Drimbarean, A., and P.F. Whelan, 2001. Experiments in colour texture analysis, *Pattern Recognition Letters*. 22: 1161-1167.
- Duda, R. O., P.E. Hart, and D.G. Stork, 2001. *Pattern Classification*, New York: John Wiley & Sons, 654p.
- Eom, K.B., 1999. Segmentation of monochrome and colour textures using moving average modeling approach, *Image and Vision Computing*, 17(3): 233-244.
- Fjortoft, R., Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, 2003. Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 41(3): 675-686.
- Forman, R.T.T., 1995. *Land Mosaics: the Ecology of Landscapes and Regions*, Cambridge:

Cambridge University Press, 652p.

Friedl, M.A., D.K. McIver, J.C.F. Hodges, X.Y. Zhang, D. Muchoney, A.H. Strahler, and

C.Schaaf, 2002. Global land cover mapping from MODIS: algorithms and early results.

Remote Sensing of Environment, 83:287-302.

Gong, P. and P.J.Howarth, 1990a. An assessment of some factors influencing multispectral

land-cover classification. *Photogrammetric Engineering & Remote Sensing*, 56: 597-603.

Gong, P. and P.J.Howarth, 1990b. Frequency-based contextual classification and gray-level

vector reduction for land-use identification. *Photogrammetric Engineering & Remote*

Sensing, 58: 423-437.

Gorte, B.H.H. and A. Stein, 1998. Bayesian classification and class area estimation of satellite

images using stratification. *IEEE Transactions on Geosciences and Remote Sensing*, 36:

803-812.

Grun, A. 1997. Automation in building reconstruction. *Photogrammetry Week'97*, Herbert

Wichmann Verlag, Heidelberg: 175-186.

Grun, A., O. Kubler, and P. Agouris, (eds.), 1995. *Automatic Extraction of Man-made Objects*

from Aerial and Space Images (II). Birkhauser Verlag, Basel.

Haala, N. and C. Brenner, 1999. Extraction of buildings and trees in urban environments.

Photogrammetric Engineering and Remote Sensing, 54: 130-137.

Healey, G., L. Wang, 1995. Illumination-invariant recognition of texture in colour images,

Journal of the Optical Society of America A, 12 (9): 1877-1883.

- Heene, G., and S. Gautama, 2000. Optimization of a coastline extraction algorithm for object-oriented matching of multisensor satellite imagery. *Proceedings of IGARSS*, Vol. 6. IEEE Press, New York, pp. 2632-2634.
- Herold, M., J. Scepan, and K.C. Clarke, 2002. The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. *Environment and Planning A*, 34: 1443 -1458.
- Hoang M. A., J.M. Geusebroek and A.W.M. Smeulders, 2005. Colour texture measurement and segmentation. *Signal Processing*, 85 (2005): 265–275.
- Hu, X., C. Vincent Tao, and B. Prenzel, 2005. Automatic segmentation of high-resolution satellite imagery by integrating texture, intensity, and colour features. *Photogrammetric Engineering & Remote Sensing*, 71 (12): 1399-1406.
- Ilgner, J.F.R., C. Palm, A.G. Schütz, K. Spitzer, M. Westhofen, and T.M. Lehmann, 2003. colour texture analysis for quantitative laryngoscopy. *Acta Otolaryngol.* 123: 730-734.
- Jahne, B., 1991. *Digital Image Processing*, New York: Springer-Verlag, 219-230.
- Jain, A., G. Healey, 1998. A multiscale representation including opponent colour features for texture recognition, *IEEE Transactions on Image Process.* 7 (1): 124-128.
- Jain, A.K., G. Healey, 1998. A multi-scale representation including opponent colour features for texture recognition, *IEEE Transactions on Image Process.* 7 (1): 124-128.
- Jensen, J.R., 2005. *Introductory Digital Image Processing: A Remote Sensing Perspective*, 3/e, Prentice Hall, Upper Saddle River, pp. 420-430.

- Jimenez, L. O., J. L. Rivera-Medina, E. Rodriguez-D z, E. Arzuaga-Cruz, and M. Ramirez-Veez, 2005. Integration of spatial and spectral information by means of unsupervised extraction and classification for homogenous objects applied to multispectral and hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43 (4):844-851.
- Kasari, M. H., J. Parkkinen, T. Jaaskelainen, R. Lenz, 1999. Multi-spectral texture segmentation based on the spectral Co-occurrence matrix. *Pattern Analysis Application*, 2(4):275–284.
- Kashyap, R. and R. Chellappa, 1983. Estimation and choice of neighbors in spatial-interaction models of images, *IEEE Transactions on Information Theory*, 29 (1): 60-72.
- Kiema, J.B.K., 2002. Texture analysis and data fusion in the extraction of topographic objects from satellite imagery. *International Journal of Remote Sensing*, 23: 767-776.
- Landgrebe, D. A., 1980. The development of a spectral-spatial classifier for earth observational data, *Pattern Recognition*, 12: 165–175.
- Lehmann, T.M., C. Palm, 2001. Local line search for illuminant estimation in real world colour scenes, *Journal of Optical Society of America A*, 18 (11): 2679-2691.
- Li, J., 2000. *Informal Settlement Modeling Using Digital Small Format Aerial Imagery*. Ph.D. Dissertation, Department of Geomatics, University of Cape Town, South Africa.
- Li, Y., 2004. *Fuzzy Similarity Measure and Application to High Resolution Colour Remote Sensing Image Processing*. M.A.Sc Thesis, Ryerson University.
- Liu, X.H., A.K. Skidmore, and H.V. Oosten, 2002. Integration of classification methods for improvement of land-cover map accuracy. *ISPRS Journal of Photogrammetry & Remote*

Sensing, 45: 257-268.

Lo, C. P. and A. K. Yeung, 2002. *Concepts and Techniques of Geographic Information Systems*, Upper Saddle River, NJ: Prentice Hall, 492p.

Lowe, D., 1995. Radial basis function networks, Arbib, M. A. (ed.), *The Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge, MA.

Lucieer, A. and A. Stein, 2002. Existential uncertainty of spatial objects segmented from remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 40: 2518-2521.

Lucieer A., A. Stein, and P. Fisher, 2005. Multivariate texture-based segmentation of remotely sensed imagery for extraction of objects and their uncertainty. *International Journal of Remote Sensing*. 26(14):2917-2936.

Mansor S., T. H. Wong and A. R. M. Shariff, 2005. Object Oriented Classification for Land Cover Mapping.

Manjunath, B. S. and R. Chellappa, 1991. Unsupervised texture segmentation using Markov random field models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13 (5): 478-482.

Marceau, D.J., P.J.Howarth, J.M.M.Dubois, and D.J.Gratton, 1990. Evaluation of the gray-level co-occurrence matrix-method for land cover classification using SPOT imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 28: 513-519.

Mayunga, S.D., 2006. *Semi-automatic Building Extraction in Informal Settlements from*

High-Resolution Satellite Imagery. Ph.D. Thesis, University of New Brunswick, pp. 37-39.

McCabe, A., T. Caelli, G. West, and A. Reeves, 2000. Theory of spatiochromatic image encoding and feature extraction, *Journal of Optical Society of America A*, 17 (10): 1744-1754.

McGraw, J.B., T.A. Warner, T. Key, and W. Lamar, 1998. Advances in high resolution remote sensing for forest ecological studies, *Trends in Ecology and Evolution*, 13(8):300-401.

McLachlan, G., and D. Peel, 2000. *Finite Mixture Models*. New York: Wiley, Wiley Series in Probability and Statistic.

Mirmehdi, M., and M. Petrou, 2000. Segmentation of colour textures, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(2):142-159.

Moody, J. and C. J. Darken, 1989. Fast learning in networks of locally-tuned processing units, *Neural Computation*, 1(2): 281-294.

Munnoz, X., J. Freixenet, X. Cufi, and J. Mart, 2003. Strategies for image segmentation combining region and boundary information, *Pattern Recognition Letters*, 24:375-392.

Ojala, T., and M. Pietik nen, 1996. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29:51-59.

Ojala, T., T.A.Ma enpa, M. Pietika Inen, J. Viertola, J. Kylo Nen, and S. Huovinen, 2002a. Outex ew framework for empirical evaluation of texture analysis algorithms. *Proceedings of 16th International Conference on Pattern Recognition*, Quebec, pp. 701-706. Available online at <http://www.outex.oulu.fi>

- Ojala, T., M. PIETIKA INEN, and A.T. MA ENPA, 2002b. Multiresolution gray-scale and rotation-invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24: 971-987.
- Palm, C., D. Keysers, T.M. Lehmann, K. Spitzer, 2000. Gabor filtering of complexhue/saturation images for colour texture classification, *Proceedings of Joint Conference on Information Science 2*, Atlantic City, NJ, USA: 45-49.
- Palm, C., T.M. Lehmann, 2002. Classification of colour textures by Gabor filtering, *Machine Graphics and Vision*, 11 (2/3): 195-219.
- Palm, C., 2004. Colour texture classification by integrative co-occurrence matrices, *Pattern Recognition Letters*, 23(7): 755-771.
- Panjwani, D.K., and G. Healey, 1995. Markov random field models for unsupervised segmentation of textured colour images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17 (10):939-954.
- Papathomas, T, R. Kashi, and A. Gorea, 1997. A human vision based computational model for chromatic texture segregation, *IEEE Transaction on Systems, Man, Cybernics*, B, 27 (3): 428-440.
- Paschos, G., 2001. Perceptually uniform colour spaces for colour texture analysis: an empirical evaluation. *IEEE Transactions on Image Process*, 10 (6): 923-937.
- Paschos, G. and M. Petrou, 2003. Histogram ratio features for colour texture classification, *Pattern Recognition Letters*, 24(1-3): 309-314.

- Poggi and A. R. P. Ragozini, 1999. Image segmentation by tree-structured Markov random field. *IEEE Signal Processessing Letters*, 7: 155-157.
- Poggi, G., G. Scarpa, and J. B. Zerubia, 2005. Supervised segmentation of remote sensing images based on a tree-structured MRF model. *IEEE Transactions on Geoscience and Remote Sensing*, 43(8): 89-94.
- Powell, M. J. D., 1987. Radial basis functions for multivariable interpolation: a review, *Algorithm for Approximation*, Mason, J. C. and M. G. Cox (Eds), 143-167, Clarendon Press, Oxford.
- Puissant, A., J. Hirsch and C. Weber, 2005. The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. *International Journal of Remote Sensing*, 26(4): 733-745.
- Qui, F., and F.S. Jensen, 2004. Opening the neural network black box and breaking the knowledge acquisition bottleneck of fuzzy systems for remote sensing image classification. *International Journal of Remote Sensing*, in press.
- Reed, T.R., J.M.H. DuBuf, 1993. A Review of recent texture segmentation and feature extraction techniques. *CVGIP Image Understanding*, 57 (3): 359-372.
- Richards, J. A., 1993. *Remote Sensing Digital Image Analysis, An Introduction*, 2nd ed., New York: Springer-Verlag.
- Rosenfeld, A. C. Ye-Wang, A. Wu, 1982. Multispectral texture, *IEEE Transaction on Systems, Man, Cybernics*, 12 (1): 79-84.

- Sanden, J.J. v. d. and D.H. Hoekman, 2005. Review of relationships between gray-tone co-occurrence, semivariance, and autocorrelation based image texture analysis approaches. *Canadian Journal of Remote Sensing*, 31(3): 207-313.
- Schenk, T., 2003. From Data to Information and Knowledge. *Digital Photogrammetry*.
- Sharma, K.M.S., and A. Sarkar, 1998. A Modified Contextual Classification Technique for Remote Sensing Data. *Photogrammetric Engineering and Remote Sensing*, 64: 997-1005.
- Song, K.Y., J. Kittler, M. Petrou, 1996. Defect detection in random colour textures, *Image Vision Computing*. 14: 667–683.
- Speis, A. and G. Healey, 1996a. Feature extraction for texture discrimination via random field models with random spatial interaction, *IEEE Transactions on Image Processing*, 5(4): 635-645.
- Speis, A. and G. Healey, 1996b. An analysis and experimental study of the performance of Markov random field applied to textured images using small samples, *IEEE Transactions on Image Processing*, 5(3): 447-458.
- Suen, P.H., G. Healey, 1999. Modeling and classifying colour textures using random fields in a random environment, *Pattern Recognition*, 32: 1009-1017.
- Swain, P.H., S. Davis, 1978. *Remote Sensing - The Quantitative Approach*, New York: McGRAW-Hill, 166-174.
- Tan, T.S.C., J. Kittler, 1994. Colour texture analysis using colour histogram, *IEEE Processing Vision Image Signal Processing*, 141 (6): 403-412.

- Thai B., G. Healey, 1998. Modeling and classifying symmetries using a multiscale opponent colour representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (11): 768-802.
- Thomas, N., C. Hendrix, and R.G. Congalton, 2003. A comparison of urban mapping methods using high-resolution digital imagery, *Photogrammetric Engineering & Remote Sensing*, 69(9): 963–972.
- Toll, D., J.R.Jensen, 1977. Detecting residential land use development at the urban fringe. *Photogrammetric Engineering & Remote Sensing*, 24:629-643.
- TopiMaenpaa, and M. Pietikainen, 2005. Classification with colour and texture: jointly or separately? *Pattern Recognition*, 37:1629-1640.
- Tran, T. N., R. Wehrens, and L. M. C. Buydens, 2005. Clustering multispectral images: A Tutorial Chemo. *Intelligence Laboratory System*, 77: 3-7.
- Tran, T. N, R. Wehrens, D. H. Hoekman, and L. M. C. Buydens, 2005. Initialization of Markov random field clustering of large remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 43(8): 321-344.
- Tso, B. and R. C. Olsen, 2005. Combining spectral and spatial information into hidden Markov models for unsupervised image classification. *International Journal of Remote Sensing*, 26, (10): 2113-2133.
- Turner, M.G., R.H. Gardner and R.V. Oeill, 2001. *Landscape Ecology in Theory and Practice: Pattern and Process*, New York: Springer-Verlag, 352p.

- Van de Wouwer, G., P. Scheunders, S. Livens, and D. Van Dyck, 1999. Wavelet correlation signatures for colour texture characterization, *Pattern Recognition*, 32(3): 443-451.
- Vincent, L., and P. Soille, 1991. Watershed in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6): 583–598.
- Wandell, B.A., 1995. *Foundations of Vision Science*, Sinauer Ass., Sunderland, MA, USA.
- Warner, T. A., and K. Steinmaus, 2005. Spatial classification of orchards and vineyards with high spatial resolution panchromatic imagery. *Photogrammetric Engineering & Remote Sensing*, 71(2): 179-187.
- Whiteside, T., W. Ahmad, 2005. A comparison of object-oriented and pixel-based classification methods for mapping land cover in northern Australia. *Proceedings of SSC2005 Spatial Intelligence, Innovation and Praxis: The national biennial Conference of the Spatial Sciences Institute*, September 2005. Melbourne: ISBN 0-9581366-2-9.
- Woods, J., 1972. Two-dimensional discrete Markovian fields, *IEEE Transactions on Information Theory*, 18: 232-240.
- Xu, Q., J. Yang and S. Y. Ding, 2005. Colour texture analysis using the wavelet-based hidden Markov model, *Pattern Recognition Letters*, 26(11): 1710-1719.
- Zadeh, L.A., 1965. Fuzzy sets. *Information and Control*, 8:338-353.
- Zhang, J., and T. Tan, 2002. Brief review of invariant texture analysis methods, *Pattern Recognition*, 35:735-737.

APPENDIX A

THE MAIN PROGRAM FOR THE PROPOSED SPECTRAL-SPATIAL INTEGRATION APPROACH

function inteseg_overlap_final

%Target image input

Cimaname=input('Object Image Name: ','s');

Cima=imread(Cimaname);

Cima=double(Cima);

[Ici,Jci,Kci]=size(Cima);

Cima1=uint8(Cima);

Cima2=Cima;

%learning image input

LC=input('image learning classes: ');

Gs=input('square group size: ');

Bns=input('The bin number: ');

pct=input('The minimum percentage for histogram peak: ');

mst=input('maximum structure number: ');

```

alp=input('Threshold coefficient: ');

morp=input('morphology box size: ');

%Segmentation and Colouring of target image

%call mpeaks and colourlibrary_w

[Vr,Vg,Vb,G]=mpeaks(Bns,pct,LC);

ColourLib=colourlibrary_w;

%multi-peak feature distribution segmentation for pixel by pixel classifying

DIST=cell(1,LC);

for m=1:Ici
    for n=1:Jci
        for i=1:LC
            for j=1:G(i)
                DIST{i}(j)=sqrt((Cima(m,n,1)-Vr{i}(j))^2+(Cima(m,n,2)-Vg{i}(j))^2+(Cima(m,n,3)-Vb{i}(j))^2);
            end
            D(i)=min(DIST{i});
        end
    end
    [d,E]=min(D);

```

```

if d<39

    if E>=1&&E<=5

        Cima(m,n,:)=ColourLib(1,:);

    elseif E>=6&&E<=9

        Cima(m,n,:)=ColourLib(2,:);

    elseif E>=10&&E<=12

        Cima(m,n,:)=ColourLib(3,:);

    elseif E==13

        Cima(m,n,:)=ColourLib(4,:);

    elseif E==14||E==15

        Cima(m,n,:)=ColourLib(5,:);

    end

else

    Cima(m,n,:)=ColourLib(6,:);

end

end

end

Cima3=uint8(Cima);

BIW=zeros(Ici,Jci);

```



```

for m=1:Ici
    for n=1:Jci
        if Cima(m,n,1)==255
            BIW(m,n)=1;                %'0' is non-building, '1' is building
        end
    end
end

end

%figuring1

BIW1=1-BIW;                %In BIW2 post-processing, '1' is non-building, '0' is building

set(0,'DefaultFigureResize','off');

figure;

imshow(Cima1),title(['Original image: ',int2str(Ici),'x',int2str(Jci),'x',int2str(Kci)],'Colour','r');

figure;

imshow(Cima3),title('pixel spectral segmented image','Colour','g');

figure;

imshow(BIW1),title('non-spatial binary building','Colour','b');

%multi-peak feature distribution segmentation for neighborhood group's classifying

%I(m,n),Q(m,n)

```

```
w=cell(1,LC);
```

```
q=cell(1,3);
```

```
M=Ici-Gs+1;
```

```
N=Jci-Gs+1;
```

```
I=ones(M,N);           %I(m,n) is each group's most tendency for a class(roof,2 or non-roof,1)
```

```
for m=1:M
```

```
    for n=1:N
```

```
        for h=1:LC
```

```
            for k=1:G(h)
```

```
                X=zeros(Gs,Gs);
```

```
                for i=1:Gs
```

```
                    for j=1:Gs
```

```
                        X(i,j)=sqrt(((Cima2((m-1)+i,(n-1)+j,1)-Vr{h}(k))^2+(Cima2((m-1)+i,(n-1)+j,2)-Vg{h}(k))^2+(C
```

```
                        ima2((m-1)+i,(n-1)+j,3)-Vb{h}(k))^2);
```

```
                    end
```

```
                end
```

```
                w{h}(k)=mean2(X);
```

```
            end
```

```
        w1(h)=min(w{h});
```

end

[w2,EE]=min(w1);

if w2<39&&EE>=10&&EE<=12 %10~12 are the roof samples

 I(m,n)=2;

end

for h=1:3

 for k=1:G(h+9) %G(h+9), Vr(h+9) denote the roof samples

 Y=zeros(Gs,Gs);

 for i=1:Gs

 for j=1:Gs

 Y(i,j)=sqrt(((Cima2((m-1)+i,(n-1)+j,1)-Vr{h+9}(k))^2+(Cima2((m-1)+i,(n-1)+j,2)-Vg{h+9}(k))^2+(Cima2((m-1)+i,(n-1)+j,3)-Vb{h+9}(k))^2);

 end

 end

 q{h}(k)=mean2(Y); %in one of K peaks, a group's average distance to a roof sample

 end

Q1(h)=min(q{h}); %among all K peaks, a group's minimum distance to a roof sample

end

$Q(m,n)=\min(Q1)$; % $Q(m,n)$ is each group's average distance to the whole roof class

end

end

%spatial context analysis based on spectral binary result

%each group spatial structure analysis

Struc=zeros(M,N);

for m=1:M

for n=1:N

for i=1:Gs

for j=1:Gs

if $BIW((m-1)+i,(n-1)+j)==1$

Struc(m,n)=Struc(m,n)+1;

end

end

end

end

end

%overlap neighborhood structure detecting

Sting=zeros(M,N);

for m=4:M-3

for n=4:N-3

if I(m,n)==1 && Struc(m,n)>0 && Struc(m,n)<=(mst+Gs-2)

for i=1:Gs

for j=1:Gs

if BIW((m-1)+i,(n-1)+j)==1 && Sting(m,n)==0

if

BIW((m-1)+i-3,(n-1)+j)==1 && BIW((m-1)+i-2,(n-1)+j)==1 && BIW((m-1)+i-1,(n-1)+j)==1

Sting(m,n)=1;

elseif

BIW((m-1)+i+1,(n-1)+j)==1 && BIW((m-1)+i+2,(n-1)+j)==1 && BIW((m-1)+i+3,(n-1)+j)==

1

Sting(m,n)=1;

elseif

BIW((m-1)+i,(n-1)+j-3)==1 && BIW((m-1)+i,(n-1)+j-2)==1 && BIW((m-1)+i,(n-1)+j-1)==1

Sting(m,n)=1;

elseif

BIW((m-1)+i,(n-1)+j+1)==1 && BIW((m-1)+i,(n-1)+j+2)==1 && BIW((m-1)+i,(n-1)+j+3)==

1

```

        Sting(m,n)=1;

    end

end

end

end

elseif I(m,n)==2&&Struc(m,n)>=(Gs*Gs-mst+Gs-3)&&Struc(m,n)<Gs*Gs

    for i=1:Gs

        for j=1:Gs

            if BIW((m-1)+i,(n-1)+j)==0&&Sting(m,n)==0

                if

BIW((m-1)+i-3,(n-1)+j)==0&&BIW((m-1)+i-2,(n-1)+j)==0&&BIW((m-1)+i-1,(n-1)+j)==0

&&BIW((m-1)+i,(n-1)+j+1)==0&&BIW((m-1)+i,(n-1)+j+2)==0&&BIW((m-1)+i,(n-1)+j+3

)==0&&BIW((m-1)+i+1,(n-1)+j)==0&&BIW((m-1)+i+2,(n-1)+j)==0&&BIW((m-1)+i+3,(n

-1)+j)==0

                    Sting(m,n)=1;

                elseif

BIW((m-1)+i+1,(n-1)+j)==0&&BIW((m-1)+i+2,(n-1)+j)==0&&BIW((m-1)+i+3,(n-1)+j)==

0&&BIW((m-1)+i,(n-1)+j-3)==0&&BIW((m-1)+i,(n-1)+j-2)==0&&BIW((m-1)+i,(n-1)+j-1)

==0&&BIW((m-1)+i-3,(n-1)+j)==0&&BIW((m-1)+i-2,(n-1)+j)==0&&BIW((m-1)+i-1,(n-1)

+j)==0

```



```

        Sting(m,n)=1;

        elseif

BIW((m-1)+i,(n-1)+j-3)==0&&BIW((m-1)+i,(n-1)+j-2)==0&&BIW((m-1)+i,(n-1)+j-1)==0
&&BIW((m-1)+i-3,(n-1)+j)==0&&BIW((m-1)+i-2,(n-1)+j)==0&&BIW((m-1)+i-1,(n-1)+j)=
=0&&BIW((m-1)+i,(n-1)+j+1)==0&&BIW((m-1)+i,(n-1)+j+2)==0&&BIW((m-1)+i,(n-1)+j
+3)==0

        Sting(m,n)=1;

        elseif

BIW((m-1)+i,(n-1)+j+1)==0&&BIW((m-1)+i,(n-1)+j+2)==0&&BIW((m-1)+i,(n-1)+j+3)==
0&&BIW((m-1)+i+1,(n-1)+j)==0&&BIW((m-1)+i+2,(n-1)+j)==0&&BIW((m-1)+i+3,(n-1)+
j)==0&&BIW((m-1)+i,(n-1)+j-3)==0&&BIW((m-1)+i,(n-1)+j-2)==0&&BIW((m-1)+i,(n-1)
+j-1)==0

        Sting(m,n)=1;

        end

        end

        end

        end

        end

        end

        end
end

```

```

%the calculation of the threshold TH for spatial homogeneity judgement

Ns=zeros(2,Gs*Gs-1);    %Ns is the total number of same structure type$tendency in the whole
    image

Qs=zeros(2,Gs*Gs-1);    %QS is the sum of group distance to 'roof' of same structure
    type$tendency in the whole image

for m=1:M
    for n=1:N
        if Struc(m,n)>0&&Struc(m,n)<Gs*Gs                %exclude groups of a pure class
            Qs(I(m,n),Struc(m,n))=Qs(I(m,n),Struc(m,n))+Q(m,n);
            Ns(I(m,n),Struc(m,n))=Ns(I(m,n),Struc(m,n))+1;
        end
    end
end

end

TH=zeros(2,Gs*Gs-1);

for u=1:2
    for v=1:Gs*Gs-1
        if Ns(u,v)>0
            TH(u,v)=Qs(u,v)/Ns(u,v);                    %calculating all the T types
        end
    end
end

```

```

        end

    end

%spatial context analysis output

for m=1:M

    for n=1:N

        if

            Sting(m,n)==0&&Struc(m,n)>=(Gs*Gs-mst+Gs-3)&&Struc(m,n)<Gs*Gs&&Q(m,n)<=TH(
                2,Struc(m,n))*alp

                for i=1:Gs

                    for j=1:Gs

                        BIW((m-1)+i,(n-1)+j)=1;

                    end

                end

            elseif

                Sting(m,n)==0&&Struc(m,n)>0&&Struc(m,n)<=(mst+Gs-2)&&Q(m,n)>TH(1,Struc(m,n))*
                1/alp)

                for i=1:Gs

                    for j=1:Gs

                        BIW((m-1)+i,(n-1)+j)=0;

```

```

        end

    end

end

end

end

%figuring2

BIW2=1-BIW;          %In BIW2 post-processing, '1' is non-building, '0' is building

figure;

imshow(BIW2),title('spatial & spectral building','Colour','r');


%morphological filtering

SE3=strel('square',morp);

J1=imclose(BIW1,SE3);

J2=imclose(BIW2,SE3);


%figuring3

figure;

imshow(J1),title('Morphological filtering of the non-spatial method','Colour','b');

figure;

```

```
imshow(J2),title('Morphological filtering of the spatial&spectral method','Colour','r');
```

```
%small object filtering
```

```
AA=J1;
```

```
BB=desmall(AA);
```

```
J3=BB;
```

```
AA=J2;
```

```
BB=desmall(AA);
```

```
J4=BB;
```

```
%figuring4
```

```
figure;
```

```
imshow(J3),title('Small object filtering of the non-spatial method','Colour','b');
```

```
figure;
```

```
imshow(J4),title('Small object filtering of the spatial&spectral method','Colour','r');
```

```
%boundary extracting and overlapping
```

```
BI=J3;
```

```
[MI,SBI]=boundary(Cima1,BI);
```

```
MI1=MI;
```

```
SBI1=SBI;
```

```

BI=J4;

[MI,SBI]=boundary(Cima1,BI);

MI2=MI;

SBI2=SBI;

%figuring5

figure;

imshow(SBI1),title('Boundary of the non-spatial method','Colour','b');

figure;

imshow(MI1),title('Overlapping of the non-spatial method','Colour','b');

figure;

imshow(SBI2),title('Boundary of the spatial&spectral method','Colour','r');

figure;

imshow(MI2),title('Overlapping of the spatial&spectral method','Colour','r');

```

APPENDIX B

THE PROGRAM OF MULTI-PEAK SUPERVISED SEGMENTATION (MSS)

```
function [Vr,Vg,Vb,G]=mpeaks(Bns,pct,LC)
```

```
%learning image input
```

```
I=cell(1,LC);
```

```
BNO=cell(1,LC);
```

```
Vb=cell(1,LC);
```

```
Vg=cell(1,LC);
```

```
Vr=cell(1,LC);
```

```
I{1}=imread('bareland1.bmp');
```

```
I{2}=imread('bareland2.bmp');
```

```
I{3}=imread('bareland3.bmp');
```

```
I{4}=imread('bareland4.bmp');
```

```
I{5}=imread('road.bmp');
```

```
I{6}=imread('lawn1.bmp');
```

```
I{7}=imread('lawn2.bmp');
```



```

I{8}=imread('lawn3.bmp');
I{9}=imread('bush.bmp');
I{10}=imread('roof1.bmp');
I{11}=imread('roof2.bmp');
I{12}=imread('roof3.bmp');
I{13}=imread('tree.bmp');
I{14}=imread('mud1.bmp');
I{15}=imread('mud2.bmp');

```

```

for i=1:LC

```

```

    CLI=I{i};

```

```

    CLI=double(CLI);

```

```

    [Ici,Jci,Kci]=size(CLI);

```

```

    %HistMatrix=zeros((Bns+Bns^2+Bns^3+5),1);    %this for 8 bins

```

```

    HistMatrix=zeros((Bns+Bns^2+Bns^3+15),1);    %this for 32 bins

```

```

    %HistMatrix=zeros((Bns+Bns^2+Bns^3+8),1);    %this for 16 bins

```

```

    for p=1:Ici

```

```

        for q=1:Jci

```

```

        Location=ceil(reshape(CLI(p,q,:),1,Kci)*Bns/256);

        P=Location(1)+Location(2)*Bns+Location(3)*Bns^2;

        HistMatrix(P+1)=HistMatrix(P+1)+1;

    end

end

HistMatrix=HistMatrix/(Ici*Jci);

II=0;

G(i)=0;

Y=pct;

for h=0:(Bns+Bns^2+Bns^3-1)

    if HistMatrix(h+1)>=Y

        Y=HistMatrix(h+1);

        II=h;

        if

            HistMatrix(i+2)<Y&&HistMatrix(i+3)<Y&&HistMatrix(i+4)<Y&&HistMatrix(i+5)<Y&&
            HistMatrix(i+6)<Y&&HistMatrix(i+7)<Y&&HistMatrix(i+8)<Y&&HistMatrix(i+9)<Y&&
            HistMatrix(i+10)<Y&&HistMatrix(i+11)<Y&&HistMatrix(i+12)<Y&&HistMatrix(i+13)<Y
            &&HistMatrix(i+14)<Y&&HistMatrix(i+15)<Y&&HistMatrix(i+16)<Y

                G(i)=G(i)+1;

```

```

        Maxbin(G(i))=II;

    end

end

if HistMatrix(h+2)<pct

    Y=pct;

end

end

for j=1:G(i)

    BNO{i}(j)=Maxbin(j);

end

end

for i=1:LC

    for j=1:G(i)

        Vb{i}(j)=(0.5+floor(BNO{i}(j)/Bns^2))*(256/Bns);

        Vg{i}(j)=(0.5+floor((BNO{i}(j)-floor(BNO{i}(j)/Bns^2)*Bns^2)/Bns))*(256/Bns);

        Vr{i}(j)=(-0.5+(BNO{i}(j)-floor((BNO{i}(j)-floor(BNO{i}(j)/Bns^2)*Bns^2)/Bns)*Bns-floor(

            BNO{i}(j)/Bns^2)*Bns^2))*(256/Bns);

    end

end

end

```

APPENDIX C

THE PROGRAM OF BUILDING BOUNDARY EXTRACTION

```
function [MI,SBI]=boundary(Cima1,BI)
```

```
MI=Cima1;
```

```
RedColour=[0,0,255];
```

```
[I,J]=size(BI);
```

```
SBI=BI;
```

```
for i=1:I
```

```
    for j=1:J
```

```
        NM=neighbormatrix(i,j,BI);
```

```
        NP=sum(NM)-NM(1);
```

```
        if (BI(i,j)==0)&((NP>0)&(NP<8))
```

```
            SBI(i,j)=0;
```

```
        else
```

```
            SBI(i,j)=1;
```

```
        end
```

```

        end

    end

    for i=1:I

        for j=1:J

            if SBI(i,j)==0

                MI(i,j,:)=RedColour;

            end

        end

    end

    end

    %figure;

    %imshow(MI);

    %figure;

    %imshow(SBI);

```