

**Urban land use extraction from
very high resolution remote
sensing images**

Mengmeng Li

Graduation committee

| | | |
|--|--|------------------------|
| Chair and Secretary | | |
| Prof.dr.ir. A. Veldkamp | | University of Twente |
| Supervisor | | |
| Prof.dr.ir. A. Stein | | University of Twente |
| Co-supervisor | | |
| Dr.ir. W. Bijker | | University of Twente |
| Members | | |
| Prof.dr.ir. M.G. Vosselman | | University of Twente |
| Prof.dr.ir. M.F.A.M. van Maarseveen | | University of Twente |
| Prof.dr. M. Herold | | Wageningen University |
| Dr.ir. K.M. de Beurs | | University of Oklahoma |

ITC dissertation number 301
ITC, P.O. Box 217, 7500 AE Enschede, The Netherlands

ISBN: 978-90-365-4328-6
DOI: 10.3990/1.9789036543286
Printed by: ITC Printing Department, Enschede, The Netherlands

© Mengmeng Li, Enschede, The Netherlands
All rights reserved. No part of this publication may be reproduced without
the prior written permission of the author.



**URBAN LAND USE EXTRACTION FROM
VERY HIGH RESOLUTION REMOTE
SENSING IMAGES**

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof.dr. T.T.M. Palstra,
on account of the decision of the graduation
committee,
to be publicly defended
on Wednesday, April 5th, 2017 at 14.45 hrs

by

Mengmeng Li
born on March 13th, 1988
in Shandong, China

This dissertation is approved by:

Prof.dr.ir. A. Stein (supervisor)

Dr.ir. W. Bijker (co-supervisor)

To my family

Summary

Rapid urbanization and population growth are challenging sustainable urban development. To deal with this, urban planners require updated land information to better understand the status of urban environments. Acquiring such information, however, is labor-intensive and time-consuming if carried out as a traditional field survey. With such survey it is also hard to get a rapid update. Remote sensing techniques serve as effective means to collect data on urban environments at various level of spatial details. Remote sensing images are therefore of great interest for extracting land use information over urban areas. The research described in this dissertation investigated and developed image analysis methods to extract urban land use information from remote sensing images. It focused on using very high resolution (VHR) satellite images with a spatial resolution below 1 m, and explored the potential of incorporating open source data.

First, an object-based method for urban land cover classification was developed. This method was targeted at two issues: obtaining an optimal image segmentation and distinguishing buildings from other man-made objects. It built a binary partition tree to hierarchically represent image segmentations. The optimal segmentation was obtained by unsupervised evaluation of segmentations using energy minimization. The method quantified the directional relationship between building and shadow objects using fuzzy sets classifying land cover at two different levels. The land cover classification was applied to Pléiades images from Wuhan, China. It performed better than a maximum likelihood classifier and a support vector machine. The directional information effectively improved the separability between building and other man-made objects.

Second, a hierarchical method based upon a binary partition tree was developed for urban road extraction. It highlighted a road region of interest, containing a set of segments, preliminarily obtained from a remote sensing image, while reducing the effects of vegetation, shadow and buildings. The road region of interest was hierarchically represented. Geometrical and structural information were considered within a region model and fused using fuzzy logic. Based on geometry, meaningful road segments were automatically extracted based upon two uncertainty related measures. Implementation was performed on a Pléiades image from Wuhan, China, and on a Quickbird image from Enschede, The Netherlands. Results showed that this method was able to group adjacent small segments of a high spectral heterogeneity

and low road-like geometrical properties. It formed meaningful road sections, and performed better than two traditionally used methods.

Third, a framework based upon a Bayesian network for urban land use extraction was developed. It started from urban land cover classification, proceeded to characterization of spatial arrangement by building types, and resulted in urban land use extraction. The spatial arrangement was characterized by quantifying the distribution of building types within a land use unit. A Bayesian network integrating the spatial arrangement and land use indicators was constructed to represent and to infer urban land use. Implementation of this framework was performed on a Pléiades image from Wuhan, China. Results showed that the spatial arrangement information significantly improved the accuracy of urban land use extraction. The Bayesian network method produced results comparable to other commonly used classifiers. This framework suited to extract urban land use from VHR images at local scale.

Fourth, a Bayesian classification of urban land use from VHR stereo images was investigated. Particular attention was paid to the incorporation of open source data for improving classification accuracy. Open Street Map data were incorporated into fuzzy decision trees based upon hierarchical Bayesian models for optimizing fuzzy rules. Scene classification results were produced by a pretrained convolution neural network with fine tuning using a UC Merced land use dataset. These were incorporated into a previously developed framework for urban land use classification. Experiments were conducted on GeoEye stereo images from Oklahoma, US. Results confirmed the transferability of the developed framework. Incorporating open source data further improved the urban land use classification. Moreover, data-driven and knowledge-driven methods were combined using Bayesian methods. This resulted into an increase in classification accuracy.

To summarize, this dissertation focused on extracting urban land use in an object-based image analysis setting, by means of analyzing urban land cover obtained from VHR images. The associated uncertainties were handled by both fuzzy sets and Bayesian methods. In this sense, this dissertation contributes to provide a framework for urban land use extraction and a set of targeted image analysis methods for extracting information from VHR remote sensing images.

Samenvatting

Snelle verstedelijking en bevolkingsgroei vormen een uitdaging voor duurzame urbane groei. Om hier mee om te gaan, hebben stedelijke planners regelmatig bijgewerkte informatie nodig om de actuele toestand van de stedelijke omgeving beter te begrijpen. Het verzamelen van dergelijke informatie is echter arbeidsintensief en tijdrovend wanneer het uitgevoerd wordt door middel van een traditionele veldstudie. Met een dergelijke studie is het ook moeilijk om een snelle actualisering te krijgen (van de gegevens). Aardobservatie technieken dienen als een effectief middel om gegevens over de stedelijke omgeving te verzamelen op verschillende niveaus van ruimtelijk detail. Aardobservatie beelden zijn daarom heel interessant om informatie te vergaren over stedelijk landgebruik. Het onderzoek beschreven in dit proefschrift onderzocht en ontwikkelde beeldanalyse technieken om informatie over stedelijk landgebruik te halen uit aardobservatie beelden. Het richtte zich vooral op zeer hoge resolutie (VHR) satellietbeelden met een ruimtelijke resolutie kleiner dan 1 meter en verkende de mogelijkheid van het betrekken van openbare data (bij de analyse).

Als eerste werd een op objecten gebaseerde methode voor de classificatie van stedelijke landbedekking ontwikkeld. Deze methode richtte zich op twee onderwerpen: het verkrijgen van een optimale beeldsegmentatie en het onderscheiden van gebouwen van andere door de mens gemaakte objecten. De methode bouwde een binaire partitie boom om beeld segmentaties op een hiërarchische manier te representeren. De optimale segmentatie werd verkregen door on-gesuperviseerde evaluatie van de segmentaties met gebruik van energie minimalisering. De methode kwantificeerde de op richting gebaseerde relatie tussen gebouw en schaduw objecten, gebruikmakend van fuzzy verzamelingen om land bedekking te classificeren op twee verschillende niveaus. De land bedekkings-classificatie werd toegepast op Pléiades beelden van Wuhan, China. Het gaf een beter resultaat dan een maximale aannemelijkheidsclassificatie en een support vector machine. De informatie gebaseerd op richting bracht een effectieve verbetering in het onderscheid tussen gebouwen en andere door de mens gemaakte objecten.

Als tweede werd een hiërarchische methode gebaseerd op een binaire partitie boom ontwikkeld voor het extraheren van urbane wegen. De methode benadrukte een aandachtsgebied voor wegen, dat een verzameling segmenten bevatte, verkregen uit een aardobservatie beeld, terwijl het de effecten van vegetatie, schaduw en gebouwen reduceerde. Het aandachtsgebied voor

wegen werd hiërarchisch gerepresenteerd. Geometrische en structurele informatie werden gewogen in een regio-model en samengevoegd met fuzzy logica. Gebaseerd op geometrie werden betekenisvolle weg-segmenten automatisch extraheerd op grond van twee op onzekerheid gebaseerde maten. De methode werd gecomplementeerd op een Pléiades beeld van Wuhan, China, en op een Quickbird beeld van Enschede, Nederland. De resultaten lieten zien dat deze methode in staat was om aangrenzende kleine segmenten met een hoge spectrale heterogeniteit en lage weg-achtige eigenschappen te groeperen. Het vormde betekenisvolle wegsecties en leverde betere resultaten dan twee traditioneel gebruikte methoden.

Als derde werd een op een Bayesiaans netwerk gebaseerd raamwerk ontwikkeld gebaseerd op een Bayesiaans netwerk voor het extraheren van stedelijk landgebruik. Het startte met de classificatie van stedelijke landbedekking en karakteriseerde vervolgens het ruimtelijke arrangement via types gebouwen en resulteerde in de extractie van stedelijk landgebruik. Het ruimtelijk patroon werd gekarakteriseerd door het kwantificeren van de verdeling van gebouw types in de landgebruikseenheid. Een Bayesiaans netwerk dat het ruimtelijke patroon en de landgebruiksindicatoren integreerde werd geconstrueerd om stedelijk landgebruik te representeren en af te leiden. Dit raamwerk werd toegepast op een Pléiades beeld van Wuhan, China. Resultaten lieten zien dat informatie over het ruimtelijk patroon significante verbetering bracht in de nauwkeurigheid van de extractie van stedelijk landgebruik. Het Bayesiaanse raamwerk produceerde resultaten vergelijkbaar met die van andere algemeen gebruikte classificaties. Dit raamwerk is geschikt om stedelijk landgebruik te extraheren uit zeer hoge resolutie (VHR) beelden op lokale schaal.

Als vierde werd een Bayesiaanse classificatie van stedelijk landgebruik onderzocht op basis van VHR stereo beelden. Speciale aandacht werd geschonken aan het incorporeren van openbaar verkrijgbare gegevens om de nauwkeurigheid van de classificatie te verhogen. Open Street Map gegevens werden verwerkt in fuzzy beslisbomen, gebaseerd op hiërarchische Bayesiaanse modellen om de fuzzy beslisregels te optimaliseren. Classificatie resultaten van beelden werden geproduceerd door een van tevoren getraind convolutie neuronaal netwerk, met een fijnregeling die gebruik maakt van een US Merced bestand van landgebruik gegevens. Dit werd verwerkt in het eerder ontwikkelde raamwerk voor de classificatie van stedelijk landgebruik. Experimenten werden uitgevoerd op GeoEye stereo beelden van Oklahoma, Verenigde Staten. Resultaten bevestigden de overdraagbaarheid van het ontwikkelde raamwerk. Het verwerken van openbaar verkrijgbare gegevens verbeterde de classificatie van stedelijk landgebruik verder. Bovendien konden, via Bayesiaanse methoden, methoden gedreven door gegevens gecombineerd worden met methoden gedreven door kennis. Dit resulteerde in een toename van de classificatie nauwkeurigheid.

Samenvattend richtte dit proefschrift zich op het extraheren van stedelijke landgebruiks-informatie in een op objecten gebaseerde beeldanalyse omgeving. Door stedelijke grondbedekking, verkregen uit VHR beelden, te analyseren. De daarmee verbonden onzekerheden werden meegenomen via zowel fuzzy verzamelingen als Bayesiaanse methoden. Op deze manier draagt dit

proefschrift bij aan het leveren van een kader voor extractie van het stedelijk landgebruik, dat gebaseerd is op een verzameling geschikte methoden voor beeldanalyse om informatie te verkrijgen uit VHR aardobservatie beelden.

Acknowledgments

When I look back the past four years, I would say that doing a PhD is full of challenges and uncertainties, but finally I have found the way to handle it. Finding this way was a long journey. I am sincerely grateful to many people and organizations for all their support to help me pass through this journey.

I would like to express my most sincere gratitude to my promoter Prof. Alfred Stein. It is a great honor to be your student. This PhD research could not have finished without your support, guidance, patience and encouragement throughout my PhD research. You trained me to be creative, critical, and enthusiastic for science. You always gave me critical and constructive feedback of my documents within an extremely short period of time. You taught me how to write English documents in a scientific way, from the usage of a single word to the structure of a PhD thesis. You even trusted my capabilities for completing a specific task much more than what I thought possible. What I learned from you will be able to benefit me throughout the rest of my life.

I am deeply grateful to my daily supervisor Dr. Wietske Bijker. Thank you so much for helping me grow from a freshman to someone who can be responsible for his research. You advised me to read the book, "How to get a PhD", so that from the very beginning, I could get to know what a PhD student should be like. You encouraged me to be more confident and brave using a poem by Erin Hanson when I proceeded to the late stage of my PhD study. You taught me how to scientifically analyze my experimental results and how to highlight research findings from them, while keeping an eye on the limitations of a specific method. Moreover, your smile makes every discussion with you successful and comfortable.

I would like to express my gratitude to Dr. Kirsten de Beurs for supporting my academic stay at the University of Oklahoma, USA. This stay greatly enriched my experience in both research and life. Particularly, you helped me to apply my research to a different urban setting in Oklahoma City, to broaden the scope of my PhD research. I am also grateful to Prof. Qingming Zhan (Wuhan University), who helped me collect data over the study area in Wuhan and supported me with the validation of some of the experimental results in this dissertation.

Many thanks to the colleagues in the EOS department of ITC. It is a great pleasure to work in such an attractive environment with a group of nice people. I am grateful to my officemates, Bashar, Caroline and Andrea, for

Acknowledgments

their kind company and help during my PhD. We have made a pretty nice and enjoyable office environment. Thanks to Teresa for all her management and consulting support, Claudio and Valentyn for their discussion on remote sensing image analysis, and Petra and George for supporting me to order remote sensing images. I further extend my gratitude to all other EOS colleagues and friends: Liang Zhou, Biao Xiong, Sudan Xu, Zhihua Xu, Yan Feng, Feifei Tang, Fashui Li, Xiaojian Wang, Zhenchao Zhang, Peng Jia, Michael Yang, Rahul, Adugna, Dewi, Ghasemi, Milad, Gustavo, Vera, Marco, Riaz, Phuong, Frank, Anand, Diogo, Zille, Shayan, and Phillipp, and many others.

I would like to sincerely thank the China Scholarship Council (CSC) for the financial support of my PhD research in ITC. I want to thank the big Chinese community in ITC/UT, and all Chinese friends I met here: Fangyuan Yu, Xiaojing Wu, Donghai Zheng, Yiwen Sun, Ying Zhang, Zhihui Wang, Linlin Li, Shaoning Lv, Jinfeng Mu, Yifang Shi, Xiaoling Wang, Yifei Xue, Xiaoxu Li, Gaoyan Wu, and many others whom I did not list here but having my sincere appreciation. Owing to their company and support, this journey has become more comfortable and enjoyable in Enschede. Thanks also go to Prof. Bob Su for sharing experience on working and living in the Netherlands. And thank you Dr. Tiejun Wang for your valuable advice on research, and many pleasant talks we had.

I would like to thank my former BSc collegemate Pengdong Zhang (University of Ghent) for his long-term company since the beginning of my PhD application, and for many discussions on research. Thanks to my former MSc colleague Dr. Liya Sun for sharing experience on PhD application. Thank you Dr. Pradeep Adhikari (now at the University of Texas at Austin) for your great help during my stay at the University of Oklahoma. You gave me a ride to super markets almost every weekend during my stay. This made my stay there much more easy. Thanks to Loes for all her management support in ITC.

I heartily indebted to my family for their constant support. They always encouraged me to follow the will of my heart, and concerned the most of my life to be healthy and happy. I owe a very special debt to my beloved wife Qian Lu, who always trusted me to be able to complete the PhD research smoothly and to lead a better life. She always takes a very good care of me, and always has my back.

*There is freedom waiting for you,
On the breezes of the sky,
And you ask "What if I fall?"
Oh but my darling,
What if you fly?*

– Erin Hanson

Contents

| | |
|---|------------|
| Summary | iii |
| Samenvatting | v |
| Acknowledgments | ix |
| Contents | xi |
| 1 Introduction | 1 |
| 1.1 Land use | 2 |
| 1.2 The need for urban land use | 2 |
| 1.3 Geospatial methods for land use extraction | 3 |
| 1.4 Image analysis methods for land use extraction from VHR images | 4 |
| 1.5 Uncertainty associated with the products derived from remote sensing images | 5 |
| 1.6 Problem statement | 7 |
| 1.7 Research objectives and questions | 8 |
| 1.8 Thesis outline | 9 |
| 2 Object-based urban land cover classification from VHR images | 11 |
| 2.1 Introduction | 13 |
| 2.2 Study area and data | 14 |
| 2.3 Methods | 14 |
| 2.4 Results | 23 |
| 2.5 Discussion | 32 |
| 2.6 Conclusions | 34 |
| 2.7 Appendix | 35 |
| 3 Urban road extraction from VHR images by a BPT | 37 |
| 3.1 Introduction | 39 |
| 3.2 Study area and data | 40 |
| 3.3 Methods | 41 |
| 3.4 Results | 48 |
| 3.5 Discussion | 52 |

xi

| | | |
|----------|---|------------|
| 3.6 | Conclusions | 54 |
| 4 | Urban land use extraction using a Bayesian network from VHR images | 57 |
| 4.1 | Introduction | 59 |
| 4.2 | Study area and data | 60 |
| 4.3 | Methods | 61 |
| 4.4 | Results | 71 |
| 4.5 | Discussion | 78 |
| 4.6 | Conclusions | 81 |
| 5 | Incorporating open source data for Bayesian classification of urban land use | 85 |
| 5.1 | Introduction | 87 |
| 5.2 | Study area and data | 88 |
| 5.3 | Methods | 89 |
| 5.4 | Results | 96 |
| 5.5 | Discussion | 104 |
| 5.6 | Conclusions | 105 |
| 6 | Synthesis | 107 |
| 6.1 | Research findings and conclusions | 108 |
| 6.2 | Reflections | 111 |
| 6.3 | Recommendations | 114 |
| | Bibliography | 115 |

List of Figures

| | | |
|------|---|----|
| 2.1 | Top panel shows the Pléiades image of the study area (true colour composite). Bottom panel shows two datasets A and B. | 15 |
| 2.2 | Workflow of the proposed classification of urban land cover. | 16 |
| 2.3 | BPT creation of an image. | 17 |
| 2.4 | Overlap between classified and reference objects. | 21 |
| 2.5 | Distribution of the spectral reflectance of six different land covers across four multi-spectral bands. | 23 |
| 2.6 | Left: the result of preliminary SVM classification in dataset A. Right: the boundaries of the classified shadow objects. | 24 |
| 2.7 | Goodness-of-fit ED , regularization EC and the total E energy values for image partitions derived from different BPTs in dataset A. (a1-a3) are energy values from BPTs using SF1-SF3 initial partitions. | 25 |
| 2.8 | Comparison of different optimal image segmentations between MIN1 and MIN2 in terms of the number of objects on datasets A (a) and B (b). | 26 |
| 2.9 | Comparison of different optimal segmentations of a piece of image in dataset A. | 28 |
| 2.10 | Top: fuzzy landscapes of shadow objects using different kernel sizes. Bottom: directional features $Drt(R_i)$ of image objects. | 29 |
| 2.11 | Comparing extracted buildings in terms of kernel sizes and ϵ_{bd} values. | 29 |
| 2.12 | Different classification results. | 30 |
| 2.13 | Distribution of the total error $TotalErr(M_i)$ index of the building objects classified by the MLC, SVM, and proposed methods on both datasets A and B. | 30 |
| 3.1 | (a) The Pléiades image (with true color composite) of Wuhan, China (study area A). (b) The Quickbird image (with true color composite) of Enschede, the Netherlands (study area B). | 40 |
| 3.2 | Workflow of the proposed road extraction method. | 41 |
| 3.3 | Tree-structure representation of a BPT sample with six initial regions. | 43 |
| 3.4 | Three MPs of pixel x_1, x_2, x_3 created by using path openings with the length L ranging from 0 to L_{max} pixels. | 44 |

List of Figures

| | | |
|------|---|-----|
| 3.5 | The obtained RoadROIs of both study areas A and B, corresponding to the top and bottom panels, respectively. | 48 |
| 3.6 | Road extraction by the proposed and other two existing methods. | 50 |
| 3.7 | Evaluation of the extracted roads, and compared with the S.V. method. | 51 |
| 4.1 | True color composite of the pansharpened Pléiades image of the study area in Wuhan, China. | 61 |
| 4.2 | The proposed Bayesian network model for extracting land uses from land covers obtained from VHR images. | 62 |
| 4.3 | Examples of different building types and land use classes in Wuhan, China, depicted by the study image. | 64 |
| 4.4 | Example of an F-Histogram. | 66 |
| 4.5 | Example of the proposed spatial signature for different pairs of image objects. | 67 |
| 4.6 | Urban land cover classified from the Pléiades VHR image. | 72 |
| 4.7 | Classified building types from the Pléiades VHR image, and visualization of the associated uncertainties. | 74 |
| 4.8 | Extracted land use from the Pléiades VHR image using the proposed Bayesian network method BN ($F_M + F_{wSA}$). | 74 |
| 4.9 | Examples demonstrating the main process of the proposed urban land use extraction from a VHR image. | 75 |
| 4.10 | Sensitivity analysis of the effect of giving different land cover classifications on final land use extractions. | 78 |
| 5.1 | Overview of study area and datasets, depicted on a GeoEye-1 image in forward view of its stereo pair. | 89 |
| 5.2 | Framework of the Bayesian classification of urban land use from VHR stereo images incorporating open source data. | 90 |
| 5.3 | Workflow of urban land cover classification. | 91 |
| 5.4 | Building types visualized on VHR images (Top) and field survey photos (Bottom). | 93 |
| 5.5 | Overview of the UC Merced land use dataset, with 21 classes Yang and Newsam (2010). The classes in bold are used in this study. | 94 |
| 5.6 | (a-d) are the derived nDSM, selected validation samples of urban land cover, classified urban land cover, and classified building types, with respect to the subset of the study image, respectively. | 97 |
| 5.7 | Urban land cover classified on the total study image. | 98 |
| 5.8 | Building types classified on the total study image. | 99 |
| 5.9 | Accuracy of the scene classification using the pretrained CNN for the UCM testing dataset. | 101 |
| 5.10 | Urban scene classification on the study image using the pretrained CNN fine tuned on UCM dataset. | 102 |
| 5.11 | Urban land use classification on the study image. | 103 |
| 5.12 | Distributions of OA and κ across ten various implementations for different land use classifications. | 103 |

List of Tables

| | | |
|------|--|----|
| 2.1 | The number of objects of various initial partitions for BPT construction in datasets A and B, respectively. | 24 |
| 2.2 | The fitted energy functions of image partitions within each BPT. | 27 |
| 2.3 | User accuracy (UA), overall accuracy (OA) and kappa coefficients of the MLC, SVM and proposed classifications for datasets A and B. Veg: vegetation, Shd: shadow, Bld: building, and Oth:other. | 31 |
| 2.4 | Confusion matrix of the proposed classification for dataset A. PA refers to Producer Accuracy. | 31 |
| 2.5 | Confusion matrix of the proposed classification for dataset B. PA refers to Producer Accuracy. | 31 |
| 2.6 | Geometric accuracies of the MLC, SVM and proposed (labelled as Pro) classifications for dataset A, evaluated by GOC, GUC and GTC global error indices, corresponding to global over-classification, under-classification and total classification errors respectively. | 32 |
| 2.7 | Geometric accuracies of the MLC, SVM and proposed (labelled as Pro) classifications for dataset B, evaluated by GOC, GUC and GTC global error indices, corresponding to global over-classification, under-classification and total classification errors respectively. | 32 |
| 2A.1 | Spectral features of image objects. | 35 |
| 2A.2 | Geometrical features of image objects. | 35 |
| 2A.3 | BPT generation procedure. | 35 |
| 3.1 | Parameter settings associated with the steps of road extraction. | 48 |
| 3.2 | Evaluation of the extracted roads by the proposed and S.V methods using completeness, correctness, and quality measures. | 52 |
| 4.1 | Features used for classifying urban land cover (Li et al., 2015a). | 63 |
| 4.2 | Features used for classifying building types. | 66 |
| 4.3 | Land use indicators used for characterizing land use. | 69 |
| 4.4 | Confusion matrix of the urban land cover classification. | 73 |
| 4.5 | Confusion matrix of the urban land use extraction by the proposed method BN ($F_M + F_{wSA}$). | 76 |
| 4.6 | Summary of the McNemar test to compare the proportions of correctly extracted land use units between BN ($F_M + F_{wSA}$) and BN (F_M). f_{12} and f_{21} are the number of discordant pairs, i.e. the number of land use units that classifiers classify differently. | 77 |

List of Tables

| | | |
|------|---|-----|
| 4.7 | Summary of the McNemar test to compare the proportions of correctly extracted land use units between BN ($F_M + F_{wSA}$) and BN ($F_M + F_{SA}$). f_{12} and f_{21} are the number of discordant pairs, i.e. the number of land use units that classifiers classify differently. | 77 |
| 4.8 | Comparison of three different urban land use extractions using the proposed BN ($F_M + F_{wSA}$), SVM ($F_M + F_{wSA}$), and RF ($F_M + F_{wSA}$) respectively. | 77 |
| 4.9 | Summary of the McNemar test of the extractions using the proposed BN ($F_M + F_{wSA}$), SVM ($F_M + F_{wSA}$), and RF ($F_M + F_{wSA}$) respectively. | 77 |
| 4A.1 | Original land use classification system for existing land use map. | 83 |
| 4A.2 | Land use classification system for this paper. | 83 |
| 5.1 | Features used for land use classification. | 95 |
| 5.2 | Confusion matrix of the urban land cover classification. | 99 |
| 5.3 | Confusion matrix of the building type classification. | 100 |
| 5.4 | Confusion matrix of the urban land use extraction. | 102 |
| 5A.1 | Features used for land cover classification. | 106 |
| 5A.2 | Features used for building type classification. | 106 |

List of Nomenclatures

Abbreviations

| | |
|-------------|---|
| BPT | Binary partition tree |
| BN | Bayesian network |
| CNN | Convolutional neural network |
| DAG | Directed acyclic graph |
| DLM | Digital landscape model |
| DSM | Digital surface model |
| DTM | Digital terrain model |
| E | Total energy |
| EC | Regularization energy |
| ED | Goodness-of-fit energy |
| EDISON | Edge detection and image segmentation |
| FENA | Fractal net evolution approach |
| F-Histogram | Histogram of forces |
| FP | False positive |
| FN | False negative |
| GIS | Geographical information system |
| GLCM | Gray-level co-occurrence matrix |
| GOC | Global over-classification error index |
| GTC | Global total error index |
| GUC | Global under-classification error index |
| HoG | Histogram of oriented gradients |
| HSI | Hue-Saturation-Intensity |
| LBP | Local binary pattern |
| LIDAR | Light detection and ranging |
| LoD1 | Level of detail 1 |
| MAP | Maximum a posteriori |
| MLC | Maximum likelihood classification |
| MP | Morphological profiles |
| MRA | Minimum region area |
| NDVI | Normalized difference vegetation index |
| nDSM | Normalized digital surface model |
| NIR | Near infrared band |
| OA | Overall accuracy |

List of Nomenclatures

| | |
|---------|---|
| OC | Over-classification error index |
| OSM | Open street map |
| OTB | Orfeo toolbox |
| PA | Producer accuracy |
| PCI | PCI geomatica software |
| RAG | Region adjacency graph |
| RBF | Radial basis function |
| RF | Random forest |
| RoadROI | Road region of interest |
| RMSE | Root-mean-square error |
| SIFT | Scale-invariant feature transform |
| SE | Structuring element |
| SF | Fractal net evolution approach segmentation |
| SM | Mean shift segmentation |
| SVM | Support vector machine |
| TC | Total error index |
| TP | True positive |
| UA | User accuracy |
| UC | Under-classification error index |
| UCM | UC Merced |
| VHR | Very high resolution |

Symbols

| | |
|-----------------------|--|
| a_i | Area of an image object |
| $\{b_1, \dots, b_S\}$ | Variables related to classified building types |
| C | Variable of land use class, $C = \{C_1, \dots, C_K\}$ |
| C_K | The states of land use class C and K is the number of land use classes |
| cm, eln | Compactness, elongation of an image object |
| d | Distance between pixels |
| $Drt(R_i)$ | Directional feature of an image object |
| $f(\cdot)$ | Function |
| F | Variables of land use features, $F = \{F_1, \dots, F_M, F_{SA}\}$ |
| $\{F_1, \dots, F_M\}$ | Variables related to land use indicators |
| F_S | Variable related to the feature from scene classification results |
| F_{SA} | Variable related to the spatial arrangement of land use |
| I | VHR image with length of N_x and width of N_y in pixels |
| H_β | Fuzzy landscape with direction angle β |
| κ | Kappa coefficient |
| k_z | Size of fuzzy structuring element |
| k_b | The number of bins in a HoG |
| len, wd | Length (i.e. major axis), width (i.e. minor axis) of an image object |
| k_c | The size of the smallest bounding box for obtaining a HoG |
| L | <i>Length</i> related to a path opening operator |

| | |
|----------------------------|---|
| L_{max} | The <i>maximum length</i> related to a path opening operator |
| LU_i | Land use unit |
| M_g, M_s | BPT region models for geometrical, spectral properties |
| $N(R_v), \coprod R_v$ | Necessity and possibility measures related to road extraction |
| $O(R_i, R_j)$ | BPT merging criteria for two neighboring objects |
| p_i | Perimeter of an image object |
| $p(\cdot)$ | Probability distribution |
| $p(\cdot \cdot)$ | Conditional probability distribution |
| P_λ | Image partition |
| R_i | Image object |
| \bar{s}_i | Mean value of the spectral bands of an image object |
| $S^{(i)}$ | The trace of $\sum^{(i)}$ |
| T | A BPT |
| v | A node of a BPT |
| w_{oa} | Weight variable related to overlapping objects |
| w_L | Weight variable related to connectivity feature derived from MP |
| w_{rm} | Weight variable related to BPT region merging |
| $\{w_1, \dots, w_S\}$ | Weight variables related to spatial arrangement characterization |
| $X_{n_i \times m_i}^{(i)}$ | Matrix of pixels of an image object R_i with n_i pixels and m_i bands |
| x_k | Random variable in BN |
| pa_k | The set of parents of x_k in BN |
| $\sum^{(i)}$ | The covariance matrix of $X_{n_i \times m_i}^{(i)}$ |
| z | Fuzzy structuring element |
| z_β, z_d | Angle-related, distance-related structuring elements |
| α, β, θ | Variables related to angles |
| δR_i | Boundary of an image object R_i |
| ϵ_{bd} | Threshold variable related to building extraction |
| ϵ_h | Threshold variable related to single-story buildings |
| ϵ_{rm} | Threshold variable related to BPT merging criteria |
| η | Unknown random variable |
| γ_{ij} | Common boundary ratio between two adjacent image objects |
| $\mu_{1,ij}$ | Membership function value related to compactness |
| $\mu_{2,ij}$ | Membership function value related to elongation |
| $\mu_{3,ij}$ | Membership function value related to HoG |
| $\mu_{4,ij}$ | Membership function value related to MP |
| $\mu_C(A, B)$ | Pairwise similarity between A and B image objects |
| ω | Solar azimuth angle |
| φ_0, φ_2 | Constant, gravitational F-histogram force |
| \mathcal{N} | Normal distribution |
| μ_0, τ_0^2 | Prior mean, variance for Bayesian analysis |
| \bar{x}, σ^2 | Sample mean, variance for Bayesian analysis |
| μ_n, τ_n^2 | Posterior mean, variance for Bayesian analysis |

Introduction

1

1.1 Land use

Land use is characterized by "the arrangements, activities and inputs by people to produce, change or maintain a certain land cover type" (FAO/UNEP, 1999). It concerns the use of land by humans, and is distinguished from land cover which refers to the biophysical properties of earth surface. Land use is directly influenced by human activities. Human activities use land for various purposes including food, settlement, energy, etc. This results into land use for agricultural, residential and commercial activities. Land use is commonly distinguished between rural and urban land use, due to variety in population density and socio-economic developments. This study focuses on urban land use. In urban areas, the land is mainly used for residential, commercial, and industrial activities, whereas agriculture and mine lands rarely occur within urban boundaries. Human activities are affected by environmental and climatic factors. These factors therefore also influence land use. An interesting example concerns the use of land by people in relation to vulnerability of natural hazards and climatic perturbations, thus affecting the decisions on land use (Anwar, 2014).

Land use information is essential to understand the interactions between humans and the environment across different spatial and temporal scales (Müller and Munroe, 2014). It is important for e.g. land management planners when analyzing the evolution of human-environment interactions. Such interactions can be linked to socio-economic factors. Such links can help to develop solutions for sustainable use of resources.

1.2 The need for urban land use

Today, 54 percent of the world's population resides in urban areas, and this figure is expected to rise to 66 percent by 2050 (UN-Habitat, 2016). Along with the rapid population growth, urban areas have also expanded remarkably. Seto et al. (2011) report an increase in global urban areas of 58,000 km² from 1970 to 2000, and estimates that the increase will be nearly tripled by 2030. Such rapid urbanization has transformed the spatial form and functionality of cities. The transformations of the spatial form can be seen from the extent, pattern and density of built-up areas, access to transportation, and connection to rural areas, while functional transformations can be elaborated upon in terms of social, economic, and environmental aspects of cities.

Urbanization brings potential opportunities to people by means of generating wealth, creating jobs, and promoting the quality of life. It is essential for fostering the prosperity of cities, and in this way has an effect on the global economy. Urbanization, however, is facing enormous challenges. Examples are the growth of slums and informal settlements, transportation, resource scarcity, inequality and insecurity, and climate change to mention a few. These challenges hinder the development of cities in a sustainable way, and potentially cause social conflicts and environmental hazards. Sustainable urban development is thus globally called for making future cities envir-

mentally sustainable and resilient, socially inclusive, safe and violence-free, and economically productive (UN-Habitat, 2016).

To achieve these goals, land use planning acts as an effective tool to address urban challenges and guide sustainable development, by means of negotiating future land and resource uses by all relevant stakeholders (GIZ, 2012). Urban land use information is fundamental for conducting land use planning over urban areas. As urbanization is proceeding at an unprecedented speed and magnitude, updated and detailed urban land use information is thus urgently needed by urban planners, decision makers and researchers in order to make urban development sustainable.

1.3 Geospatial methods for land use extraction

Advancements in geospatial techniques, including Geographical Information Systems (GIS), remote sensing, database system, spatial statistics, and geovisualization, have greatly promoted the development of the way that land use data are collected, stored, managed, analyzed and visualized (Thakur et al., 2012). Land management and planning departments have used geospatial techniques to derive and manage diverse thematic maps related to land cover and land use. Nowadays, land-related maps have been derived from multisource geospatial data, including remote sensing images, field survey observations and other ancillary data, whereas GIS has been used as an effective tool for management and visualization.

Urban planners require thematic and topological urban land information at fine spatial details to plan cities at local scale. Traditionally, extraction of such land information is mainly based upon field survey, which is costly for large areas, in particular for a whole nation. The development of remote sensing technologies has changed the traditional way of land information extraction. Aerial images were initially used to obtain land use information over urban areas. This was done by human interpretation, and was constrained to specific areas due to limited accessibility of aerial images. Since the launch of IKONOS satellite in 1999 which provides remote sensing images with 1 m spatial resolution in the panchromatic channel, increasing attention has been paid to the use of very high resolution (VHR) satellite images for urban land mapping. Over the last decade, the number of VHR satellites has been growing rapidly. This provides a huge volume of available VHR images; meanwhile, the price of VHR images has been decreasing sharply.

Very high resolution (VHR) satellite images provide landscape information at an increasing level of spatial detail, the latest VHR satellites Worldview-3 and -4 are able to offer remote sensing images with spatial resolution of 0.3 m. Existing studies have indicated the benefits of using VHR images for urban-related studies such as urban growth analysis and urban population estimation (Pozzi and Small, 2005; Taubenböck et al., 2012; Patino and Duque, 2013). So far, many studies have been conducted on obtaining urban *land cover* information from VHR images, by exploiting the use of the spectral, textual and geometrical information of objects (Myint et al., 2011).

Less research has been done on extracting *land use* information from VHR images. Extracting urban land use from VHR images is more challenging than land cover extraction, because (1) VHR images directly record the spectral reflectance of earth surface, and (2) human activities cannot be directly inferred by interpreting the tone, texture or shapes of image features. Therefore, advanced image analysis methods incorporating expert knowledge or ancillary data (if available), should be investigated for land use extraction using VHR images.

1.4 Image analysis methods for land use extraction from VHR images

A primary step of a land use classification system is to obtain basic land use units. A land use unit has homogenous land use, and may comprise several land cover types. Land use extraction from medium- and low- resolution remote sensing images, is commonly based on image pixels (Wharton, 1982; Letourneau et al., 2012). It directly assigns pixels into various land use classes according to the spectral, textual and spatial characteristics of pixels. Use of spatial information has been emphasized for land use extraction (Ünsalan and Boyer, 2011). A straightforward practice to derive spatial information for pixel-based methods is to use a moving window for image neighborhood definition (Eyton, 1993). Spatial information is thus derived within this regular neighborhood. The limitations are evidently seen: (1) the optimal neighborhood is hard to determine, although improvements have been investigated for example by van der Kwast et al. (2011); (2) such a regular neighborhood does not correspond to a real neighborhood on the ground, particular in VHR images; (3) the classification results tend to smooth the boundaries between land use classes (Barnsley and Barr, 1997; Herold et al., 2003; Lackner and Conway, 2008). Pixel-based methods, based upon regular neighborhoods, are thus less used for land use extraction from VHR images.

Below is given an overview of existing relevant studies on land use extraction from VHR images, based upon how land use units are defined and how spatial information is used for land use classification.

1.4.1 Vector/Parcel-based methods

Vector/Parcel-based methods rely on existing boundary data, such as parcels and road networks, to obtain homogenous land use units (Wu et al., 2006; Lackner and Conway, 2008; Wu et al., 2009a; Hu and Wang, 2013; Alahmadi et al., 2015). Within a land use unit, spatial information can thus be derived. Land use indicators that mainly consider the coverage ratio and density of land cover features are commonly used (Herold et al., 2003; Hu and Wang, 2013; Lowry and Lowry, 2014; Novack et al., 2014). The advantage of using such methods is that land use units can be well defined, whereas a major concern is that existing boundary data for specifying land use units might not be available for some areas.

1.4.2 Graph-based methods

Graph-based methods facilitate object-based land use extraction from VHR images. A graph-based method starts from the construction of a planar graph, in which the spatial information can be calculated within a sub-graph. Such a graph can represent the spatial relations between land cover objects, where a sub-graph corresponds to a basic land use unit. Interesting work on using graph-based methods for land use classification is given by Barnsley and Barr (1997); Comber et al. (2012) and Walde et al. (2014). The difficulty, however, is to effectively partition a graph into meaningful sub-graphs.

1.4.3 Scene-based methods

In recent years, studies on urban scene classification have drawn increasing attention to extracting semantic information from remote sensing images, in particular from VHR images (Yang and Newsam, 2010; Vaduva et al., 2013; Castelluccio et al., 2015; Zhao et al., 2016). Similar to natural scene classification in computer vision, an urban scene refers to a subset of an image, and is defined according to its semantic interpretation, e.g. an airport, a parking lot, or a sparse residential area. Yang and Newsam (2010) conducted a scene-based land use classification using bag-of-visual-words, and published the UC Merced (UCM) land use dataset, which is now frequently used. Similar studies can also be found in Cheriyyadat (2014); Chen and Tian (2014). Currently, interesting studies have used deep learning methods for scene-based land use classification (Castelluccio et al., 2015; Jean et al., 2016). These methods, particularly based upon convolutional neural networks (CNNs) (Krizhevsky et al., 2012), performed better than bag-of-visual-words. Scene-based methods usually characterize spatial information based upon low-level image descriptors, e.g. scale-invariant feature transform (SIFT) in Cheriyyadat (2014) and local binary patterns (LBP) in Li et al. (2015b).

1.5 Uncertainty associated with the products derived from remote sensing images

Error and uncertainty arise at various stages of the production of maps from remotely sensed imagery (Comber et al., 2008). Many studies have indicated the importance of their quantification as well as of the uncertainty of the products derived from those images (Castilla and Hay, 2007; Stein et al., 2009; Comber et al., 2010; Stehman and Wickham, 2011; Wickham et al., 2013). Shi (2009) categorized uncertainty into three types: positional, attribute, and topological uncertainty. Each type of uncertainty can be further divided into error, vagueness and ambiguity (Fisher, 1999). This dissertation in particular focuses on the positional and attribute uncertainty. Positional uncertainty is defined as the differences between the measured position of a spatial feature and its actual value on the ground. Attribute uncertainty is defined as the closeness of an attribute value to its actual

1. Introduction

value. For an urban land use classification system, key uncertainties include: (1) the classification of land cover, (2) the definition of land use classes, (3) the definition and delineation of land use units, (4) the quantification of the discriminating criteria of land use units, and (5) the final land use classification.

Land cover classes can be well-defined, for example by standard land cover classification systems. Then the boundaries between objects with different land cover classes can be delineated with a relatively high certainty (Wu et al., 2009b). The main uncertainty refers then to positional and attribute errors (Comber et al., 2010). Positional errors in this context are defined as the degree of compliance with which the coordinates of points determined from a map differ from the coordinates determined by surveys or other independent procedures that are generally accepted as accurate (ASPRS and ASCE, 1994). In this case, positional errors are mainly caused by mis-registration and geometric corrections (Castilla and Hay, 2007). Their quantity is evaluated by means of e.g. the Root Mean Square Error (RMSE) (Congalton and Green, 2009). The attribute error of land cover is inversely related to the classification accuracy of land cover. For its quantification, a classification accuracy assessment is a useful tool, commonly resulting in a confusion matrix. Pixel-based accuracy assessments by the confusion matrix, however, do not provide the distribution of uncertainty in space: we know the degree of uncertainty, but not where areas of high and low accuracies are. Increasing attention has been paid to object-based assessments (Lucieer and Stein, 2002; Clinton et al., 2010; Persello and Bruzzone, 2010).

The definition and delineation of land use units is crucial in urban land use classification systems. It is however hard to automatically obtain the explicit boundaries at a high precision for every land use unit based upon a VHR image, as transition zones may exist between different land use types. In an urban environment, for example, residential, commercial and industrial are either difficult to delineate, they may exist as mixed land use types or gradual changes from one class to the other may occur.

Urban land use classification is implemented based upon a set of discriminating criteria. Since urban land use units are characterized by means of an abstraction and approximation of the real world (Shi, 2009), uncertainty is involved in characterizing land use units and quantification of the discriminating criteria. This uncertainty also affects classification results, as a classifier uses this information as the input. In addition, such uncertainty can propagate towards derived products, like planning activities and economic evaluations, and an uncertain input produces an uncertain output: it is 'garbage in = garbage out' (Heuvelink et al., 1989). Combining and integrating uncertainties collected during the main steps in creating products from remote sensing images, remains an important topic in remote sensing and is still insufficiently addressed (Stein et al., 2009).

1.6 Problem statement

Although several image analysis methods exist for urban land use extraction from VHR images, challenges are still faced to implement a (semi-) automatic land use extraction while achieve a satisfactory accuracy. The challenges are augmented in urban areas. The current practice to extract detailed land use information for urban planners, still relies on human interpretation. The main problems that hinder an effective land use extraction from VHR images by means of (semi-) automatic image analysis methods, are the following.

1. Complex urban environments

Urban environments are generally dominated by built-up areas, where most of human activities carried out. Urban built-up areas consist of buildings, squares, parking lots, roads, and other man-made structures, and are used for various land use purposes. Particularly, buildings are essential for residential, commercial and industrial use. Extracting building information in detail is thus important for urban land use extraction. In urban areas, buildings, however, are constructed with diverse shapes and sizes. Unlike natural land cover classes such as vegetation and water, buildings usually have similar color, tone, and texture as other man-made objects. Moreover, tall trees and shadow effects may also hinder building extraction from VHR images. This makes building extraction from VHR images intrinsically difficult, and consequently causes difficulties for land use extraction. This also increases difficulties in dealing with the uncertainty related to the classification of land cover (including building extraction) in complex urban areas, as discussed in the section 1.5.

2. Issues related to VHR images

Besides potential opportunities offered by VHR images for urban applications, issues and problems related to VHR images have also been aware by researchers. Gamba et al. (2011) listed four types of issues related to VHR images: geometrical problems, spectral problems, mapping limits and challenges, and change detection. Geometrical problems refer to the geometric accuracy, need for orthorectified VHR images, and challenges of analyzing multi-view VHR images. The geometrical problems refer to the positional uncertainty of using VHR images. A main spectral problem is the relatively coarse spectral resolution of current VHR images compared with their fine spatial resolution. This results in difficulties in urban land information extraction. Although many studies have investigated object-based methods for land information extraction, obtaining the optimal image segmentation of a VHR is still a challenge. In contrast, poorly defined image objects can increase the uncertainty of land information extracted from VHR images in aspects of vagueness and ambiguity (Fisher, 1999).

3. Land use intraclass heterogeneity and interclass similarity

In urban environments, intraclass heterogeneity and interclass similarity affect land use extraction from VHR images, particularly over ill-planned urban areas. As is well-known, a land use unit may consist

of various land cover types. Land use indicators, based upon the coverage ratio and density of land cover features, have been widely used for land use extraction from low- and medium- resolution remote sensing images. They, however, fail to effectively characterize different land use types from VHR images at the local level. For example, both high-density residential and commercial land use types are covered largely with buildings and some vegetation on a VHR image. Commercial use may also be confused with industrial use. In addition, land use units with a same type might be diversely distributed in feature space. These problems correspond to the uncertainty issue of quantifying the discriminating criteria of land use units. Therefore, it is crucial to explore an effective way to characterize land use over urban areas.

4. Land use varies between different regions

Land use varies between different regions and countries because of different planning strategies and environmental conditions across time. There is no single land use system meeting all possible urban applications. This corresponds to the uncertainty issue with respect to the definition of land use classes. The fact is that land use classes need to be well tuned based upon a specific area. Developing land use extraction methods that are robust to different urban settings from VHR images and flexible to incorporate ancillary data, is thus of great significance, but challenging.

1.7 Research objectives and questions

The main objective of this dissertation is to develop and apply advanced image analysis methods for extracting land use information from VHR remote sensing images over urban areas. To this aim, the following specific objectives and questions are addressed.

1. Develop a method to extract land cover over complex urban environments from VHR remote sensing images.

Research question: *Can the optimal image segmentation be obtained based upon an binary partition tree and energy measures?*

Research question: *Can the directional relationship between shadow and building objects be quantified to improve building extraction?*

2. Develop a method to extract urban road from VHR remote sensing images.

Research question: *Can binary partition trees be used to effectively represent and extract urban roads?*

3. Develop a method to infer land use from the land cover derived from VHR remote sensing images in urban areas.

Research question: *Can spatial arrangement be effectively characterized to improve urban land use extraction?*

4. Apply the method developed in objective 3 to a different study area using VHR stereo images.

Research question: *Can the method developed in objective 3 be applicable to a different urban area?*

Research question: *Can open source data be successfully incorporated to improve urban land use extraction?*

1.8 Thesis outline

This thesis is a compilation of six chapters. Besides the Introduction and Synthesis, the four remaining chapters are based upon papers that have been published in, or submitted to, ISI journals, and are independently structured as Abstract, Introduction, Study area and data, Methods, Experiments, Discussion, and Conclusions. The six chapters are described as follows:

- **Chapter 1** gives the general introduction of this thesis. It highlights the importance of urban land use information for sustainable urban development, and illustrates the potential of using VHR remote sensing images to extract urban land use. Research objectives and questions are then motivated. Subsequent chapters are targeted at these objectives and address the corresponding questions.
- **Chapter 2** presents an object-based method for urban land cover classification from VHR images. An image segmentation based upon binary partition tree optimized by energy measures is presented. Building extraction based upon the directional relationship between shadow and building objects is also presented.
- **Chapter 3** presents a hierarchical method based upon a binary partition tree for urban road extraction from VHR images. The potential of using binary partition tree for region-based road extraction is investigated. Fuzzy logic is used to fuse different types of road features.
- **Chapter 4** presents a framework based upon a Bayesian network for urban land use extraction from VHR images. Spatial arrangement is characterized by quantifying the distribution of building types within a land use unit. The spatial arrangement information and commonly used land use indicators are integrated by a Bayesian network. Land use extraction is conducted over Wuhan study area.
- **Chapter 5** investigates the incorporation of open source data into a Bayesian classification of urban land use from Very High Resolution (VHR) stereo satellite images. Land use extraction is conducted over Oklahoma study area.
- **Chapter 6** summarizes the results obtained from Chapters 2 - 5, and answers the research questions in Chapter 1. Reflections and recommendations of this thesis are also given.

Use of Binary Partition Tree and energy minimization for object-based classification of urban land cover

2

This chapter is based on the published paper: Li, M., Bijker, W., Stein, A., 2015. Use of Binary Partition Tree and energy minimization for object-based classification of urban land cover. *ISPRS Journal of Photogrammetry and Remote Sensing* 102, 48–61.

Abstract

Two main challenges are faced when classifying urban land cover from very high resolution satellite images: obtaining an optimal image segmentation and distinguishing buildings from other man-made objects. For optimal segmentation, this work proposes a hierarchical representation of an image by means of a Binary Partition Tree (BPT) and an unsupervised evaluation of image segmentations by energy minimization. For building extraction, we apply fuzzy sets to create a fuzzy landscape of shadows which in turn involves a two-step procedure. The first step is a preliminary image classification at a fine segmentation level to generate vegetation and shadow information. The second step models the directional relationship between building and shadow objects to extract building information at the optimal segmentation level. We conducted the experiments on two datasets of Pléiades images from Wuhan City, China. To demonstrate its performance, the proposed classification is compared at the optimal segmentation level with Maximum Likelihood Classification and Support Vector Machine classification. The results showed that the proposed classification produced the highest overall accuracies and kappa coefficients, and the smallest over-classification and under-classification geometric errors. We conclude first that integrating BPT with energy minimization offers an effective means for image segmentation. Second, we conclude that the directional relationship between building and shadow objects represented by a fuzzy landscape is important for building extraction.

Keywords: Urban land cover, Binary Partition Tree, Energy minimization, Fuzzy landscapes, Directional relationship

2.1 Introduction

Very high resolution (VHR) satellite images provide landscape information at an increasing level of spatial detail. Many studies have tested the benefits of using VHR images for urban-related studies such as urban growth analysis and urban population estimation (Pozzi and Small, 2005; Patino and Duque, 2013; Taubenböck et al., 2012). A high spatial resolution of VHR images, however, also results in a decrease in spectral separability. Image objects, which can provide a vast number of spectral, textural, geometrical and contextual features, have been preferably used for classifying VHR images instead of using individual pixels (Blaschke et al., 2014). It is therefore challenging to generate the optimal segmentation with the specified level of detail for representing meaningful objects (Hamedianfar et al., 2014; Tian and Chen, 2007).

An effective approach for representing image objects is to use a Binary Partition Tree (BPT) (Salembier and Garrido, 2000). A BPT stores a set of meaningful image objects in a tree structure, where the hierarchy of nodes corresponds to the decreasing (or increasing) level of detail of image objects (Kurtz et al., 2012). It has been effectively used for processing both natural scenes and satellite images. For example, Vilaplana et al. (2008) and Liu et al. (2011) utilized a BPT for object detection from colour images. In remote sensing, Valero et al. (2013) applied a region-based BPT for representing and classifying hyperspectral images. Alonso-Gonzalez et al. (2013) used a BPT for region-based representation of both synthetic aperture radar (SAR) and hyperspectral images. Kurtz et al. (2012) investigated the use of BPT for extracting complex urban patterns from high resolution satellite images.

A BPT represents a hierarchy of image objects as well as a series of hierarchical image partitions. BPT-based image segmentation aims to identify a set of meaningful image objects with respect to a specified degree of homogeneity. The methods for pruning a BPT are often performed in a supervised way either using cut examples (Kurtz et al., 2012) or by giving cut thresholds (Alonso-Gonzalez et al., 2013; Valero et al., 2013). In the literature, energy minimization has been used for unsupervised evaluation of image segmentations. For example, the region-oriented scale-sets theory proposed by Guigues et al. (2006) was built upon the energy minimization principle by minimizing a two-term-based energy. Philipp-Foliguet and Guigues (2008) investigated an unsupervised evaluation approach for image segmentation based on energy minimization.

In urban areas, buildings are generally of diverse shapes, sizes and materials, and similar to other man-made objects such as parking lots, squares and roads (Niemeyer et al., 2014). Therefore it is difficult to distinguish buildings from other man-made objects. In a satellite image, a shadow object adjacent to a building object is casted in a specific direction. The directional relationship between buildings and shadows therefore provides important supplementary information for image classification, particularly in discriminating buildings from other man-made objects that do not cast shadows. To deal with the uncertainty associated with the directional relationship,

several studies have proposed to use fuzzy sets to create a fuzzy landscape of a reference object for modelling this directional relationship (Krishnapuram et al., 1993; Bloch, 1999; Akçay and Aksoy, 2010; Ok et al., 2013).

This paper aims to deal with obtaining the optimal image segmentation and extracting buildings from other man-made objects. Specifically, we apply Binary Partition Tree and energy minimization for image segmentation, and exploit fuzzy sets to create a fuzzy landscape of shadows for building extraction. The remainder of this paper is organized as follows: Section 2.2 describes the study area and data. Section 2.3 starts with describing the framework of the proposed method, followed by illustration of each component of the proposed method. Section 2.4 presents experimental results and related analysis, followed by discussion in Section 2.5 and the conclusions of this research in Section 2.6.

2.2 Study area and data

In this study we used a Pléiades image that was recorded over a central region of the city of Wuhan, China, on 11 July 2013. This image was acquired with a solar azimuth angle of 140.8° , and has a 2m spatial resolution in the multispectral bands, Blue (430–550 nm), Green (490–610 nm), Red (600–720 nm) and near infrared-NIR (750–950 nm), and a 0.5m spatial resolution in the panchromatic-Pan band. This 0.5m Pan band is resampled from the 0.7m raw panchromatic band by the image provider using spline resampling. The study area contains various urban land use types (residential, commercial, industrial and undeveloped regions) and land cover types (building, road, vegetation, and bare soil). For this work, we selected two different datasets representing typical urban land forms to apply the proposed method (Figure 2.1). Dataset A with size 591×633 pixels and B with size 576×618 pixels are located in a residential area and a school respectively, and both involve the challenge of image segmentation and building extraction. Particularly, the buildings in dataset A exhibit diverse shapes and sizes, leading to poor characterization by geometrical features. Dataset B contains man-made objects that are made of different materials, resulting in high spectral heterogeneity. To improve the performance of image segmentation, two image pre-processing steps were conducted: Gram-Schmidt pan-sharpening, and 3×3 image median filtering. The reference datasets used for accuracy assessment were digitized from the datasets A and B.

2.3 Methods

The proposed approach for classifying urban land cover is conducted at two image segmentation levels, namely initial and optimal segmentations, and two main procedures, namely preliminary land cover classification and building extraction (Figure 2.2). Preliminary land cover classification serves to extract vegetation and shadow objects at the first segmentation level (i.e. an over segmentation), referring to Figures 2.2(b and e). The next step



Figure 2.1: Top panel shows the Pléiades image of the study area (true colour composite). Bottom panel shows two datasets A and B.

extracts building objects at the second segmentation level (i.e. an optimal segmentation), referring to Figures 2.2(d and i). Specifically, a BPT is used to hierarchically represent a VHR image and create a hierarchy of image segmentations (Figure 2.2(c)). Energy minimization is used to identify the optimal image segmentation among a set of possible segmentations (Figure 2.2(d)). Fuzzy sets are used to create a fuzzy landscape for representing the directional relationship between building and shadow objects for building extraction (Figures 2.2(f, g and h)). Finally, we overlay the preliminarily classified land cover and the extracted buildings to obtain the final classification result (Figure 2.2(j)). The classification result is assessed by both pixel-based and object-based methods shown as Figure 2.2(k). In the following sections, we will describe the proposed method in detail.

2. Object-based urban land cover classification from VHR images

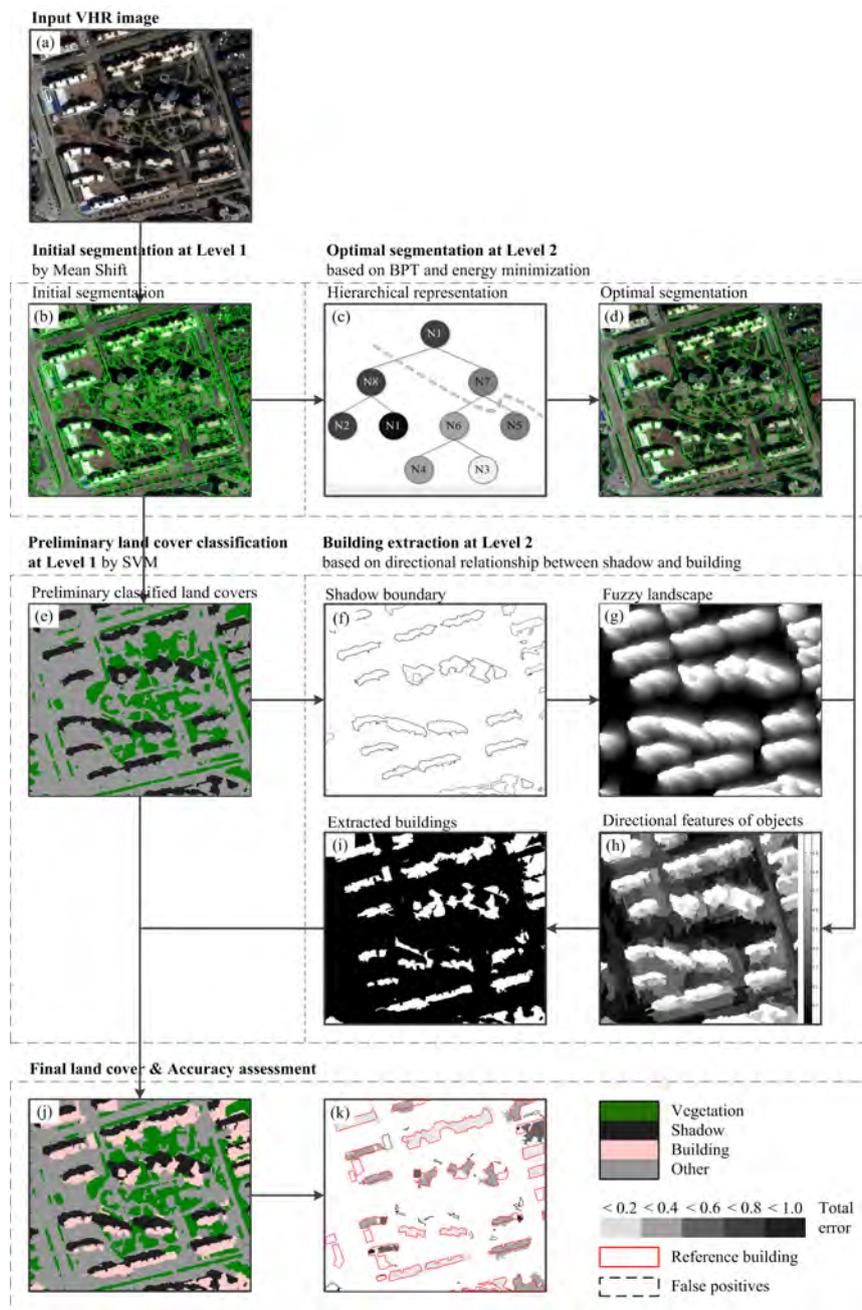


Figure 2.2: Workflow of the proposed classification of urban land cover. The proposed method refers to two segmentation levels - initial and optimal segmentation levels at which the preliminary land cover classification and building extraction are carried out respectively.

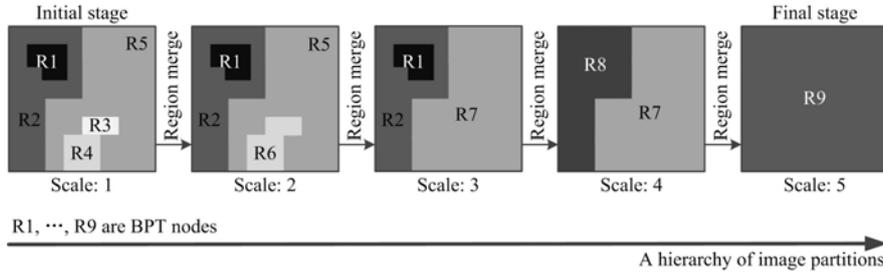


Figure 2.3: BPT creation of an image. It starts with an initial stage that represents an image partition of the finest detail, then iteratively merges two adjacent regions to create new partitions of decreasing detail, and ends with a final stage that forms a single region that represents a partition of the coarsest detail.

2.3.1 Preliminary land cover classification

The proposed method starts from an initial partition of a VHR image (Figure 2.2). We use the Mean Shift algorithm to produce an over segmentation to be the initial segmentation (Comaniciu and Meer, 2002). A Support Vector Machine (SVM) is then used to classify vegetation and shadow objects based on spectral features (Table 2A.1 in Appendix) at this segmentation level. We use the radial basis function for the SVM classifier, and a one-versus-one strategy to deal with multi-class classification (Hsu et al., 2010). The values for penalty and kernel parameters of SVM are determined via grid search algorithm and 5-fold cross validation (Huang and Wang, 2006).

2.3.2 Optimal segmentation

2.3.2.1 BPT representation

We consider an image I with size $N_I = N_x \times N_y$, where N_x and N_y are the image length and width in pixels, respectively. Let $P_1 = \{R_1^{(1)}, R_2^{(1)}, \dots, R_M^{(1)}\}$ be an initial partition of I , where image regions $R_i^{(1)}$ and $R_j^{(1)}$ satisfies $R_i^{(1)} \cap R_j^{(1)} = \emptyset$, $i \neq j$ and $\bigcup_{i=1}^M R_i^{(1)} = I$. The BPT of I can be created by the pair-wise region merging of $M - 1$ steps. This BPT hierarchically stores $2M - 1$ nodes including M leaves and one root according to the decreasing level of detail of image regions. Then the set of image partitions equals $\{P_\lambda\}$ with $\lambda \in \{1, 2, \dots, M\}$, where $P_\lambda = \bigcup_{i=1}^{M-\lambda+1} R_i^{(\lambda)}$ is an image partition after $\lambda - 1$ region mergings, and P_1 and P_M refer to image partitions at the initial and final stages of this BPT (Figure 2.3). In this way, the parameter λ behaves as a scale parameter (Guigues et al., 2006). For shorthand, the superscript (λ) will be dropped for following description. The BPT creation involves two key components: the region model and region merging criterion.

In this work, the initial partition for a BPT creation relies on an initial image segmentation that is an over segmentation case produced from an existing segmentation method. The region model specifies how regions are

2. Object-based urban land cover classification from VHR images

represented and how to compute the model of the union of two regions (Kurtz et al., 2012). It is thus defined as a two low-level process, based on spectral and geometrical properties, $M_s(R_i)$ and $M_g(R_i)$, respectively, with

$$M_s(R_i) = \bar{s}_i, \quad (2.1)$$

$$M_g(R_i) = a_i. \quad (2.2)$$

Here \bar{s}_i and a_i are the mean value of the spectral bands and the area size of R_i . Furthermore, the region model of the two merged regions is defined as

$$M_s(R_{ij}) = \frac{\bar{s}_i \times a_i + \bar{s}_j \times a_j}{a_i + a_j}, \quad (2.3)$$

$$M_g(R_{ij}) = a_i + a_j, \quad (2.4)$$

where $M_s(R_{ij})$ and $M_g(R_{ij})$ are the spectral and geometrical parts of the union of R_i and R_j .

The region merging criterion defines the dissimilarity of adjacent regions and the order in which regions are going to be merged (Vilaplana et al., 2008), adopting a region merging criterion proposed by Liu et al. (2011). The criterion measures the spectral difference, area property and adjacency degree of two adjacent regions. Let $M_s(R_i)$ and $M_g(R_i)$ be the spectral and geometrical models of region R_i respectively, and let δR_i and δR_j be the boundaries of R_i and R_j respectively; the common boundary between R_i and R_j is referred as to δR_{ij} . Suppose that the degree of adjacency between two adjacent regions R_i and R_j is quantified by a common boundary ratio γ_{ij} :

$$\gamma_{ij} = \max\left(\frac{\delta R_{ij}}{\delta R_i}, \frac{\delta R_{ij}}{\delta R_j}\right). \quad (2.5)$$

Then the merging criterion $O(R_i, R_j)$ of two adjacent regions is defined as

$$O(R_i, R_j) = (1 - \gamma_{ij})(M_g(R_i) \cdot |M_s(R_i) - M_s(R_{ij})| + M_g(R_j) \cdot |M_s(R_j) - M_s(R_{ij})|). \quad (2.6)$$

This approach is justified because it first merges two adjacent regions of smaller $O(R_i, R_j)$ with a large common boundary, whereas $O(R_i, R_j)$ with a small common boundary are merged later. It is a simple and transparent method. The reader can refer to Table 2A.3 in Appendix for the implementation of a BPT.

2.3.2.2 Energy minimization

Following the creation of an image BPT, a set of image partitions $\{P_\lambda\}$ with $\lambda \in \{1, 2, \dots, M\}$ emerged. We select a meaningful image partition P_{λ^*}

with $P_{\lambda^*} \in \{P_{\lambda}\}$ to represent the optimal segmentation. The selection is conducted by evaluating the quality of all possible image partitions. The evaluation of an image partition is conducted in an unsupervised way based on energy minimization. The optimal segmentation thus equals an image partition that reflects the minimal energy according to used energy measures. We adopt two popular energy measures to evaluate the quality of image partitions, namely goodness-of-fit and regularization energy.

The goodness-of-fit energy is fully described in Mumford and Shah (1989), and Philipp-Foliguet and Guigues (2008). Obtaining the optimal segmentation $P_{\lambda^*} = \{R_i\}$ is considered as modelling of a piece-wise smooth function $f(x, y)$ with $\{(x, y) \mid 1 \leq x \leq N_x, 1 \leq y \leq N_y\}$, where a region R_i corresponds to a piece of relevant function f_i . Then the goodness-of-fit energy of $ED(R_i)$ is quantified based on the Euclidean distance between region R_i and its corresponding model f_i . Let $X_{n_i \times m_i}^{(i)}$ be the matrix of pixels of R_i , where n_i and m_i are the number of pixels and spectral bands respectively. Considering spectral bands as random variables, we can calculate the covariance matrix $\sum^{(i)}$ of $X_{n_i \times m_i}^{(i)}$. Let $S^{(i)}$ be the trace of $\sum^{(i)}$. The goodness-of-fit energy $ED(R_i)$ is thus defined as

$$ED(R_i) = n_i \times S^{(i)}. \quad (2.7)$$

The regularization energy is used to penalize a segmentation that has ragged boundaries of objects. The regularization energy $EC(R_i)$ of region R_i is defined based on the perimeter δR_i of R_i as

$$EC(R_i) = \delta R_i. \quad (2.8)$$

Combining $ED(R_i)$ with $EC(R_i)$, the total energy E_{λ} of an image partition P_{λ} is defined as

$$E_{\lambda} = \frac{1}{c_1} \sum_{R_i \in P_{\lambda}} ED(R_i) + \frac{1}{c_2} \sum_{R_i \in P_{\lambda}} EC(R_i), \quad (2.9)$$

where c_1 and c_2 are the normalization coefficients of the goodness-of-fit and regularization energies respectively; c_1 equals to $\max(\sum_{R_i \in P_{\lambda}} ED(R_i))$; c_2 is set to the size of the image, equals to $N_x \times N_y$ (Philipp-Foliguet and Guigues, 2008).

To identify an image partition that reflects the optimal segmentation, we apply a straightforward (labelled as MIN1) strategy that identifies the partition that has the lowest total energy globally among a set of possible partitions. In addition, a data fitting (labelled as MIN2) strategy is used for comparison of MIN1. It first fits the total energies of possible partitions into a function, then identifies a partition that has the lowest energy according to this fitted function.

2.3.3 Building extraction based on the directional relationship between shadow and building

Given a reference (i.e. shadow) object R_{ref} and a direction with angle β , a fuzzy landscape $H_{\beta}(R_{ref})$ of an image I specifies a membership function

2. Object-based urban land cover classification from VHR images

to the reference object R_{ref} , taking values between 0 and 1. It specifies the degree of satisfying the directional relationship to R_{ref} (Bloch, 1999). In this work, the membership value $H_\beta(R_{ref})(x)$ is quantified by the angle $\theta_\beta(x, b)$ and distance $d(x, b)$, where pixel $x \in I$ and $b \in R_{ref}$. The angle $\theta_\beta(x, b)$ is the angle between the vector from x to b and the unit vector along the direction β with respect to the horizontal axis, whereas $d(x, b)$ is the Euclidean distance between x and b . We use fuzzy morphological dilation to generate the fuzzy landscape. To have an efficient computation, we use boundary pixels δR_{ref} of R_{ref} , instead of using all pixels, for the fuzzy dilation operation (Ok et al., 2013). The fuzzy landscape $H_\beta(R_{ref})(x)$ is calculated as

$$H_\beta(R_{ref})(x) = (\delta R_{ref} \oplus z)(x) \cap R_{ref}^c, \quad (2.10)$$

where \oplus is the dilation operator, and R_{ref}^c is the complement of R_{ref} . The z is a non-flat fuzzy structuring element (SE), composed of an angle-related SE z_β and a distance-related SE z_d , for dilation, with

$$z = z_\beta(x) \times z_d(x), \quad (2.11)$$

$$z_\beta(x) = \max\left(0, 1 - \frac{2}{\pi}\theta_\beta(x, o)\right), \quad (2.12)$$

$$z_d(x) = \max\left(0, 1 - \frac{2 \times d(x, o)}{\sqrt{2} \times k_z}\right). \quad (2.13)$$

Here o is the centre pixel of z_β , $\theta_\beta(x, o)$ is the angle between the vector from x to o and the vector along the direction β , $d(x, o)$ is the distance between x and o , and k_z is the size of z_d . Furthermore, the supplementary metadata file provides the image data with information about solar azimuth angle ω that can be used to infer the direction angle β with respect to the horizontal axis, i.e. $\beta = \omega - \frac{\pi}{2}$ (Ok et al., 2013).

Given fuzzy landscape $H_\beta(R_{ref})$ and optimal segmentation P_λ , the directional relationship between buildings and shadows can be transformed into directional features of objects by averaging the pixel values of $H_\beta(R_{ref})(x)$ within each object R_i . The directional feature $Drt(R_i)$ is obtained as

$$Drt(R_i) = \frac{1}{n_i} \sum_{p=1}^n H_\beta(R_{ref})(x_p), \quad (2.14)$$

where n_i is the number of pixels of R_i , and x_p indicates a pixel of R_i . The building extraction is performed based on $Drt(R_i)$ using a threshold value ϵ_{bd} ,

$$\begin{cases} 1, & Drt(R_i) \geq \epsilon_{bd}, & R_i \neq R_{ref}, \\ 0, & Drt(R_i) < \epsilon_{bd}, & R_i \neq R_{ref}. \end{cases} \quad (2.15)$$

where 1 indicates object R_i is building object and 0 otherwise.

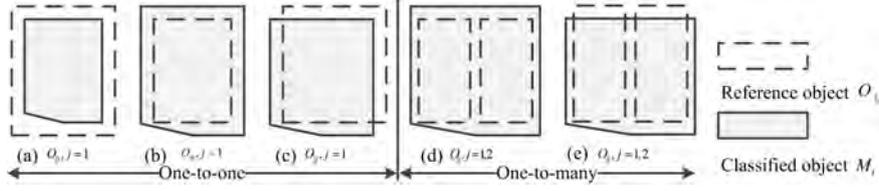


Figure 2.4: Overlap between classified and reference objects. One-to-one instances of over-classification (a), under-classification (b), and both over- and under-classification (c). One-to-many instances of under-classification (d), and both over- and under-classification (e).

2.3.4 Accuracy assessment

The extracted buildings are overlaid with the results of preliminary land cover classification to obtain the final classification results. To validate the proposed classification, we adopt a pixel-based and an object-based methods for accuracy assessment. The pixel-based assessment is based on the overall accuracy and kappa coefficient computed from confusion matrix. We have modified the commonly used under-segmentation and over-segmentation geometric error indices to perform object-based assessment for classification results (Clinton et al., 2010; Persello and Bruzzone, 2010).

Suppose that a classified object M_i overlaps a set of reference objects O_{ij} with $j = 1, 2, \dots, r$, where $r = 1$ and $r > 1$ refer to one-to-one and one-to-many instances respectively (Figure 2.4). Let $\text{area}(M_i)$ and $\text{area}(O_{ij})$ be the area of M_i and O_{ij} respectively, and $\text{area}(M_i \cap O_{ij})$ be the overlapping area of the pair of objects (M_i, O_{ij}) . For one-to-many instances, the commonly used over-segmentation and under-segmentation error indices (e.g. Clinton et al. (2010)) can lead to under estimation of the classification accuracy for one-to-many instances (Figures 2.4(d and e)). This paper proposes two modified over-classification and under-classification error indices by introducing a weight variable for each pair of overlapping objects (M_i, O_{ij}) . The weight variable is evaluated by a size (area) ratio between a reference object $\text{area}(O_{ij})$ and the total reference objects $\sum_{j=1}^r \text{area}(O_{ij})$. The modified over-classification $\text{OC}(M_i)$, and under-classification $\text{UC}(M_i)$ error indices are defined as

$$\text{OC}(M_i) = \sum_{j=1}^r \left(w_{oa} \cdot \left(1 - \frac{\text{area}(M_i \cap O_{ij})}{\text{area}(O_{ij})} \right) \right), \quad (2.16)$$

$$w_{oa} = \frac{\text{area}(O_{ij})}{\sum_{j=1}^r \text{area}(O_{ij})},$$

$$\text{UC}(M_i) = 1 - \frac{\sum_{j=1}^r \text{area}(M_i \cap O_{ij})}{\text{area}(M_i)}. \quad (2.17)$$

2. Object-based urban land cover classification from VHR images

Combining $OC(M_i)$ with $UC(M_i)$, the total error $TC(M_i)$ index is defined as

$$TC(M_i) = \sqrt{\frac{OC(M_i)^2 + UC(M_i)^2}{2}}. \quad (2.18)$$

Furthermore, we use $OC(M_i)$, $UC(M_i)$ and $TC(M_i)$ indices to obtain three global error indices: GOC, GUC, and GTC, for estimating the global properties of the classification results. These three global error indices are defined as

$$GOC = \sum_{i=1}^m \left(OC(M_i) \cdot \frac{\text{area}(M_i)}{\sum_i^m \text{area}(M_i)} \right), \quad (2.19)$$

$$GUC = \sum_{i=1}^m \left(UC(M_i) \cdot \frac{\text{area}(M_i)}{\sum_i^m \text{area}(M_i)} \right), \quad (2.20)$$

$$GTC = \sum_{i=1}^m \left(TC(M_i) \cdot \frac{\text{area}(M_i)}{\sum_i^m \text{area}(M_i)} \right), \quad (2.21)$$

where m is the number of classified objects with respect to a specified class type in a classification map.

2.3.5 Performance comparison

Based upon the optimal segmentation derived from a BPT and energy minimization, the commonly used maximum likelihood classification (MLC) and SVM classification are implemented to demonstrate the performance of the proposed method. Here the algorithm settings for the SVM classifier are same as those in the preliminary SVM classification of the proposed method. We calculate seven spectral object features and eight geometrical objects features for both MLC and SVM (Tables 2A.1 and 2A.2 in Appendix). Furthermore, we distinguish between Fractal Net Evolution Approach (FNEA) segmentation (SF) and Mean Shift segmentation (SM) for BPT creation. The SF segmentation is implemented by eCognition software (Benz et al., 2004). For the SF segmentation, we take shape and compactness parameters equal to 0.1 and 0.5 respectively (by default), whereas we distinguish three scale parameters equal to 40, 45 and 50, labelling them as SF1, SF2 and SF3 respectively. The SM segmentation is implemented by Edge Detection and Image Segmentation (EDISON) system (Comaniciu and Meer, 2002). For the SM segmentation, we take spatial band width and range band width parameters equal to 7 and 6.5 respectively (by default), whereas we distinguish three minimum region area parameters equal to 10, 15 and 20, labelling them as SM1, SM2 and SM3 respectively.

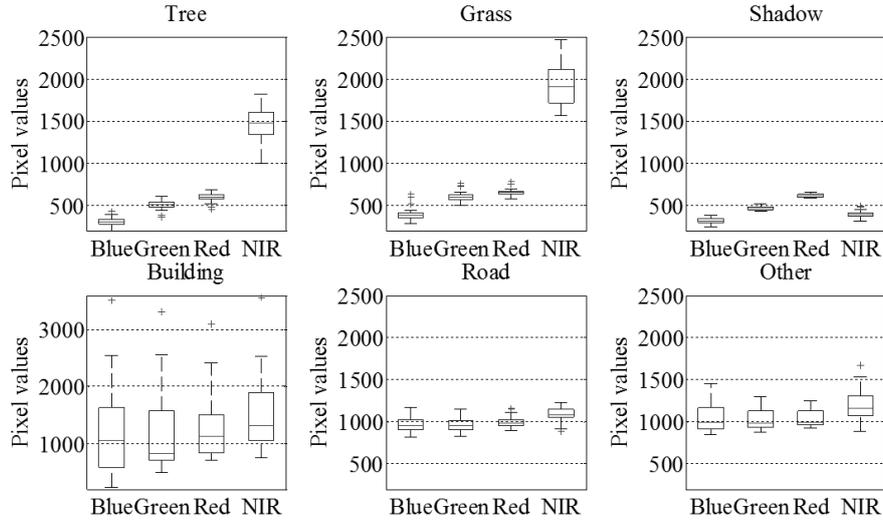


Figure 2.5: Distribution of the spectral reflectance of six different land covers across four multi-spectral bands.

2.4 Results

2.4.1 Preliminary land cover classification

We analysed the spectral response over six main urban land covers: tree, grass, shadow, building, road and other land cover (Figure 2.5). Figure 2.5 shows that the distribution of the tree reflectance varies between different bands, which is close to that of grass, but quite different from those of building, road and other land cover. Reflectance of both tree and grass in NIR band is evidently higher than in blue, green and red bands. Reflectances of shadow in the blue and NIR bands are lower than other two bands. In feature space, the building, road and other land covers are highly overlapping. To summarize, tree, grass and shadow can be classified based on spectral features, but building, road and other cannot be distinguished purely on spectral features.

The initial segmentation for preliminary land cover classification was created by the Mean Shift algorithm. The values for spatial band width and range band width were by default set to 7 and 6.5 respectively. The minimum region area (MRA) parameter was set to 15. We then used SVM to classify the initial image objects into vegetation, shadow and other classes, based on seven spectral features (Table 2A.1 in Appendix). For SVM training dataset, we selected 30, 24 and 30 samples that had good representation of real objects for vegetation, shadow and other classes respectively. Figure 2.6 shows the result of the preliminary land cover classification on dataset A. This figure also displays boundaries of the classified shadow objects that were used for the following fuzzy landscape generation.

2. Object-based urban land cover classification from VHR images

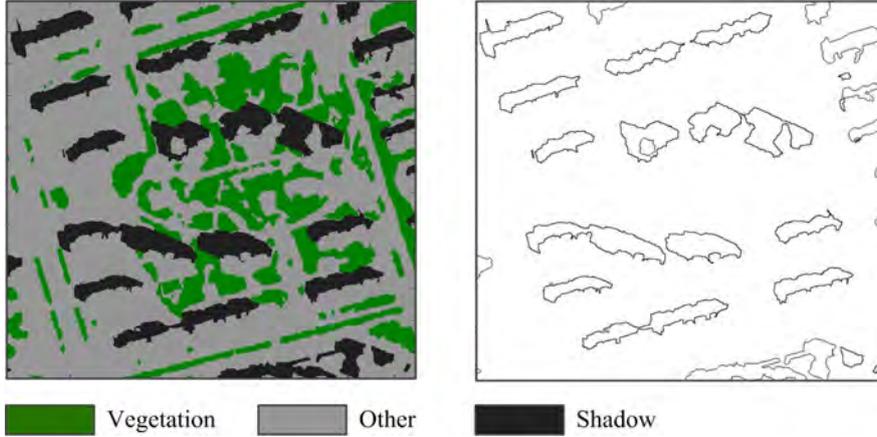


Figure 2.6: Left: the result of preliminary SVM classification in dataset A. Right: the boundaries of the classified shadow objects.

Table 2.1: The number of objects of various initial partitions for BPT construction in datasets A and B, respectively. MRA refers to the parameter of minimum region area of Mean Shift algorithm.

| Initial partitions | Dataset A | | Dataset B | |
|--------------------|------------|-------------------|------------|-------------------|
| | Parameters | Number of objects | Parameters | Number of objects |
| SF1 | Scale = 40 | 3849 | Scale = 40 | 3234 |
| SF2 | Scale = 45 | 3156 | Scale = 45 | 2658 |
| SF3 | Scale = 50 | 2639 | Scale = 50 | 2217 |
| SM1 | MRA = 10 | 3796 | MRA = 10 | 3686 |
| SM2 | MRA = 15 | 3198 | MRA = 15 | 3119 |
| SM3 | MRA = 20 | 2819 | MRA = 20 | 2763 |

2.4.2 Optimal segmentation

2.4.2.1 BPT representation

To observe the effect of initial image partitions on BPT generations, we generated various initial partitions using FNEA and Mean Shift methods for BPT construction on both datasets A and B. The number of objects with respect to different initial partitions is given in Table 2.1.

2.4.2.2 Energy minimization

We calculated goodness-of-fit energy ED , regularization energy EC and the total energy E , for each set of image segmentations hierarchically represented by a BPT. Figure 2.7 plots the curves of these energies for hierarchical image segmentations on dataset A. To keep displaying consistence, we took 2000 hierarchical image segmentations for each BPT. Figure 2.7 shows that ED has an exponential increase with the rise of scales. In contrast, EC exhibits an exponential decrease. The total energy E first shows a slight decrease, then it goes into a trough, and ends with a rapid increase.

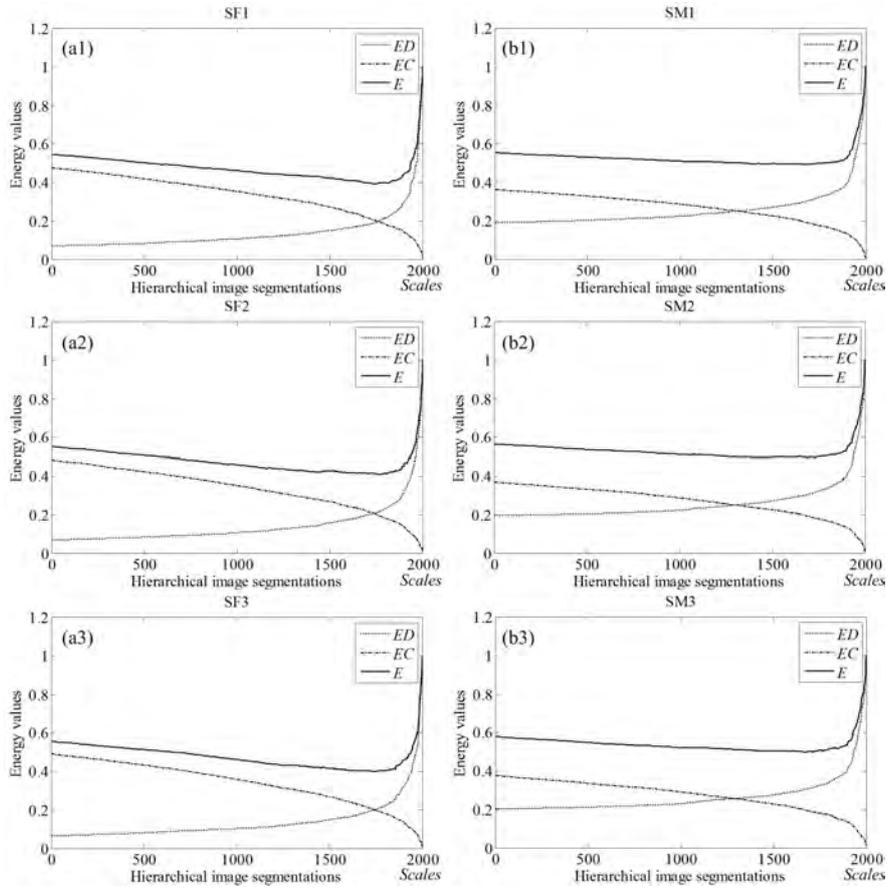


Figure 2.7: Goodness-of-fit ED , regularization EC and the total E energy values for image partitions derived from different BPTs in dataset A. (a1-a3) are energy values from BPTs using SF1-SF3 initial partitions. (b1-b3) are energy values from BPT using SM1-SM3 initial partitions. For an individual graph, the horizontal axis indicates the hierarchal image partitions with an increase of scale parameter, the vertical axis denotes the energy values.

As illustrated in Section 2.3.2.2, two strategies, MIN1 (i.e. straightforward) and MIN2 (i.e. data fitting) were used to obtain the optimal segmentation. For MIN2, we fitted the total energy E of a segmentation into an exponential function with respect to the number of objects m against the value of E . To keep data fitting consistency, we again took 2000 hierarchical image segmentations for each BPT on both datasets A and B. Table 2.2 lists the details of the fitting functions. It shows that all coefficients of determination R^2 exceed 0.9. We compared the number of objects for the obtained optimal image segmentations between MIN1 and MIN2 (Figure 2.8). We noticed that the number of objects of the optimal segmentations produced by MIN1, (squares), was larger than those of MIN2, (triangles). Furthermore,

2. Object-based urban land cover classification from VHR images

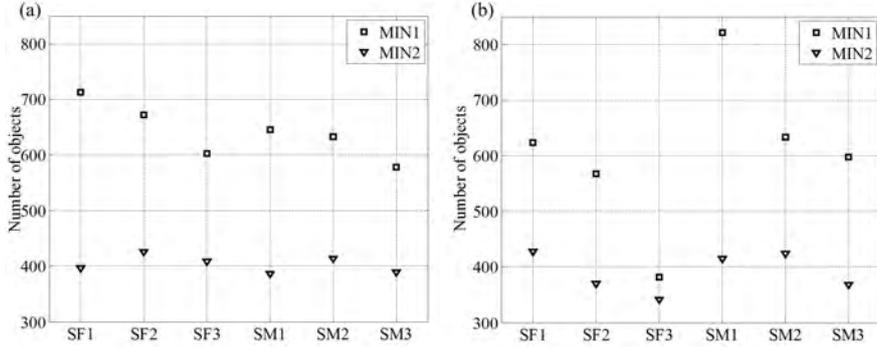


Figure 2.8: Comparison of different optimal image segmentations between MIN1 and MIN2 in terms of the number of objects on datasets A (a) and B (b). SF1 refers to the optimal segmentation from the BPT using SF1 initial partition.

Figure 2.8 shows that the number of objects of the optimal segmentations from MIN1 varies more than those from MIN2.

We visually compared the results of optimal segmentations between MIN1 and MIN2 (Figure 2.9). Comparing different initial segmentation methods, e.g. Figure 2.9(a) vs. Figure 2.9(c), we noticed that Mean Shift produced fewer objects and delineated better contours than FNEA. Comparing different energy minimization strategies, e.g. Figure 2.9(a) vs. Figure 2.9(b), we can see more segments using MIN1 than using MIN2. As indicated by red boxes in Figures 2.9(c and d), however, MIN2 produced a slight under-segmentation. For example, a vegetation object was merged into its neighbouring man-made object. To avoid under-segmentation, we chose MIN1 strategy to obtain the optimal segmentations (i.e. SM2_MIN1) on datasets A and B for following experiments.

2.4.3 Building extraction based on the directional relationship between shadow and building

Figure 2.10 displays the fuzzy landscape of the classified shadow objects as well as the directional features $Drt(R_i)$ of image objects at the optimal image segmentation level. Figures 2.10(a-c) show different fuzzy landscapes by varying the kernel size of the fuzzy structuring element, and the corresponding feature maps $Drt(R_i)$ of image objects are given in Figures 2.10(d-f). Building extraction was performed on the directional features $Drt(R_i)$ given a specific ϵ_{bd} value. Figure 2.11 compares the extracted buildings between different ϵ_{bd} values. Figures 2.11(a2 and b2) show a better performance than the other. Specifically, as indicated by blue boxes in Figures 2.11(a1, b2, c1 and c2), the extracted buildings was large mixed with the surroundings. By contrast, the buildings were insufficiently extracted in the yellow boxes of Figures 2.11(a3, b3 and c3). In this work, we used a kernel size $k_z = 100$ and $\epsilon_{bd} = 0.5$ for building extraction, because a large value of k_z would lead

Table 2.2: The fitted energy functions of image partitions within each BPT. The independent variable m and dependent variable E respectively refer to the number of objects of and the total energy of an image partition represented by a BPT.

| Datasets | Fitted energy function | R^2 | RMSE | Minimum value |
|----------|--|--------|--------|---------------|
| A | SF1 $E = 0.6658 \times e^{-0.01227m} + 0.1233 \times e^{0.0004265m}$ | 0.9308 | 0.0224 | 0.1511 |
| | SF2 $E = 0.6185 \times e^{-0.00982m} + 0.1228 \times e^{0.0005848m}$ | 0.9536 | 0.0199 | 0.1670 |
| | SF3 $E = 0.5884 \times e^{-0.00955m} + 0.1283 \times e^{0.0006663m}$ | 0.9549 | 0.0221 | 0.1803 |
| | SM1 $E = 0.6513 \times e^{-0.01367m} + 0.1506 \times e^{0.0002676m}$ | 0.9481 | 0.0179 | 0.1703 |
| | SM2 $E = 0.6474 \times e^{-0.01136m} + 0.1363 \times e^{0.0004131m}$ | 0.9545 | 0.0180 | 0.1676 |
| | SM3 $E = 0.6459 \times e^{-0.01179m} + 0.1430 \times e^{0.0004550m}$ | 0.9464 | 0.0199 | 0.1773 |
| B | SF1 $E = 0.6152 \times e^{-0.00978m} + 0.1903 \times e^{0.0004042m}$ | 0.9737 | 0.0140 | 0.2357 |
| | SF2 $E = 0.6131 \times e^{-0.01115m} + 0.1989 \times e^{0.0004673m}$ | 0.9791 | 0.0136 | 0.2463 |
| | SF3 $E = 0.6146 \times e^{-0.01183m} + 0.2136 \times e^{0.0005020m}$ | 0.9879 | 0.0116 | 0.2644 |
| | SM1 $E = 0.6155 \times e^{-0.01158m} + 0.2456 \times e^{0.0002178m}$ | 0.9299 | 0.0209 | 0.2739 |
| | SM2 $E = 0.5706 \times e^{-0.00988m} + 0.2399 \times e^{0.0003110m}$ | 0.9360 | 0.0201 | 0.2824 |
| | SM3 $E = 0.5922 \times e^{-0.01162m} + 0.2574 \times e^{0.0003293m}$ | 0.9642 | 0.0155 | 0.2988 |

to more computation time.

2.4.4 Accuracy assessment

We implemented the MLC, SVM and proposed classifications for both datasets A and B. For MLC and SVM, the classifications were based on the obtained optimal segmentations by means of a BPT and energy minimization. The training dataset for MLC and SVM classifications were the same, and contained 40 objects for each of these four classes: vegetation, shadow, building, and other, respectively. We selected sample objects that had good representation of real objects. Seven spectral and eight geometrical features of image objects were calculated for MLC and SVM classifications (Tables 2A.1 and 2A.2 in Appendix). The results of these three classifications are shown in Figure 2.12. This figure shows that more building objects were extracted by the proposed method than by MLC and SVM. Those produced an evident misclassification between building and other classes, as the building and other man-made objects show a significant overlap in spectral reflectance. In this case, the shadow provides useful supplementary information for distinguishing buildings from other man-made objects. Moreover, comparing Figure 2.12(a4) with Figures 2.12(a2 and a3), the vegetation extracted by the proposed method shows a finer level of object detail. Some building objects, indicated by red boxes in Figure 2.12, however, failed to be extracted by the proposed method. These buildings lacked a shadow object, because as a result of the direction and situation of the building rel-

2. Object-based urban land cover classification from VHR images

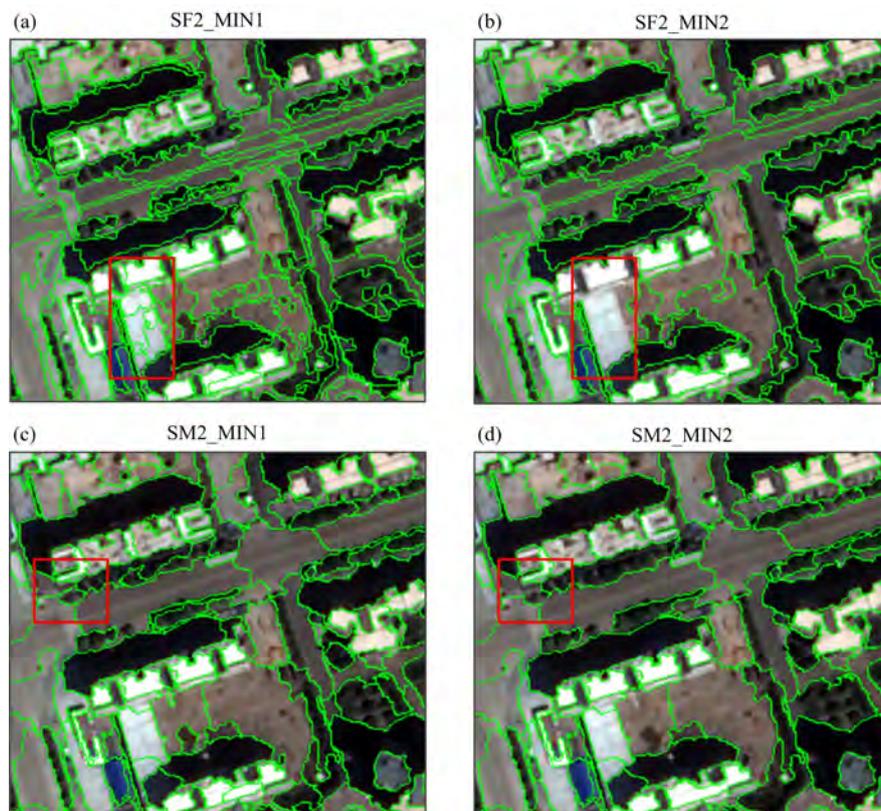


Figure 2.9: Comparison of different optimal segmentations of a piece of image in dataset A. (a) and (b) are optimal segmentations produced by MIN1 and MIN2, respectively, using SF2 initial partitions. (c) and (d) are optimal segmentations produced by MIN1 and MIN2, respectively, using SM2 initial partitions.

ative to the illumination and other buildings, the shadows of these buildings were mainly cast onto themselves or on other buildings. Therefore, in these cases building detection using the proposed method failed, while MLC still detected the buildings.

We assessed the results of the MLC, SVM and proposed classifications based on confusion matrices using pixel-by-pixel comparison between classification results and reference data. The results of accuracy assessments in terms of user accuracy (UA), overall accuracy (OA) and kappa coefficient are given in Table 2.3, whereas Tables 2.3 and 2.4 give confusion matrices for the proposed classifications on datasets A and B respectively. Table 2.3 shows that, as compared with MLC and SVM, the proposed method yielded the highest OA and kappa accuracies for both datasets A and B, equal to 0.83 and 0.76, and 0.76 and 0.66 respectively. There is no significant difference between MLC and SVM classifications on datasets A or B. The SVM clas-

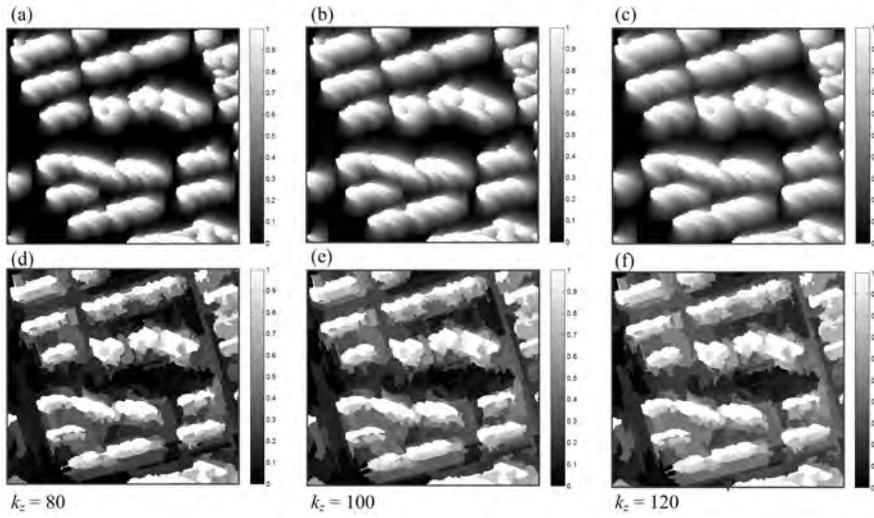


Figure 2.10: Top: fuzzy landscapes of shadow objects using different kernel sizes of (a) $k_z = 80$, (b) $k_z = 100$, and (c) $k_z = 120$. Bottom: directional features $Drt(R_i)$ of image objects, where a larger value of $Drt(R_i)$ indicates a stronger directional relationship between this object to the corresponding shadow object.

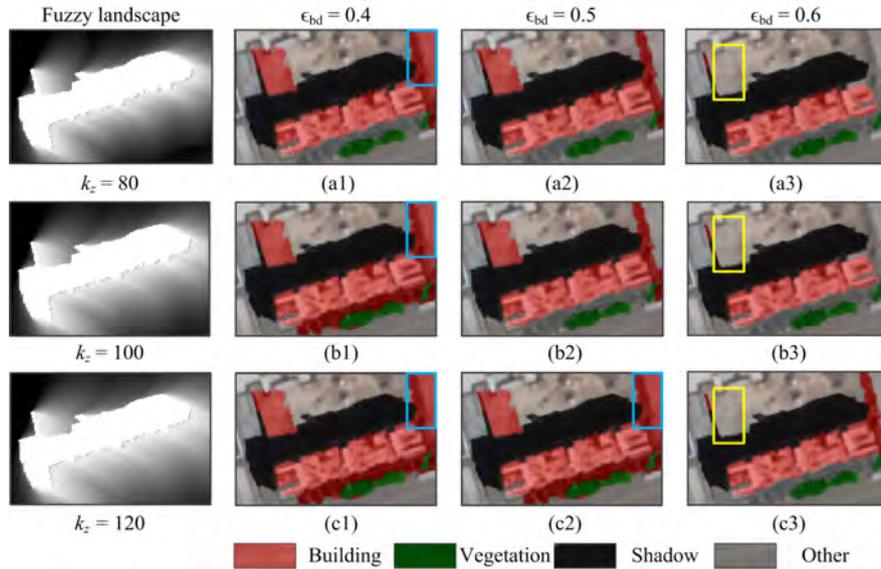


Figure 2.11: Rows (A-C) compare the difference of extracted buildings in terms of different kernel sizes, i.e. $k_z = 80$, 100 and 120. Columns (e.g. a1-a3) compare the difference of extracted buildings in terms of different ϵ_{bd} values, i.e. $\epsilon_{bd} = 0.4$, 0.5 and 0.6.

2. Object-based urban land cover classification from VHR images

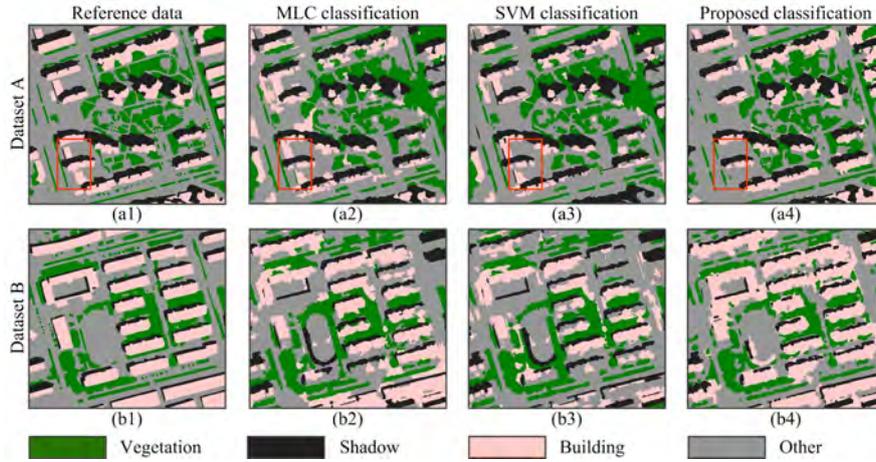


Figure 2.12: Different classification results. Top: classifications from the MLC, SVM and proposed methods for dataset A as well as reference data. Bottom: classifications from the MLC, SVM and proposed methods for dataset B as well as reference data.

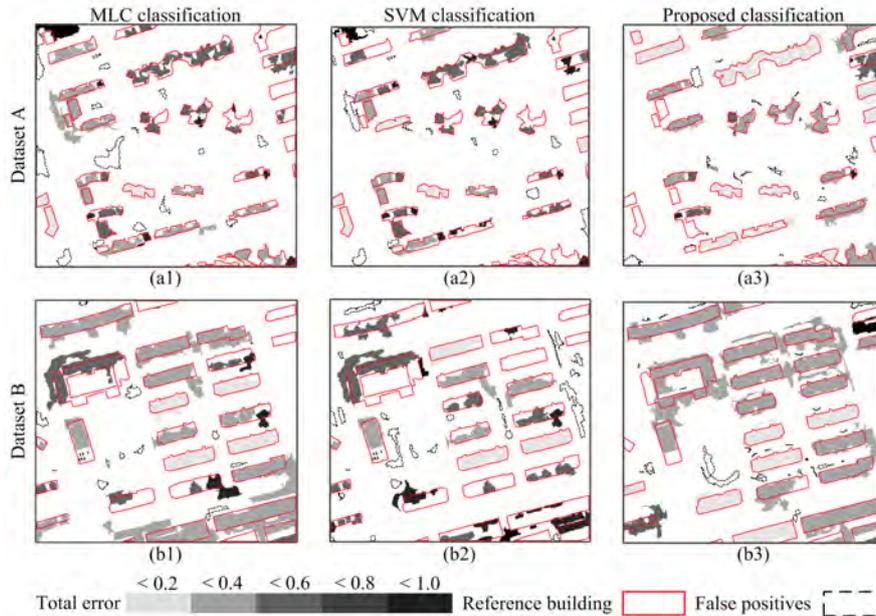


Figure 2.13: Distribution of the total error $TotalErr(M_i)$ index of the building objects classified by the MLC, SVM, and proposed methods on both datasets A and B. False positives refer to the extracted building objects that do not overlap with the reference objects.

sification produced a slightly higher OA and kappa accuracy than MLC on

Table 2.3: User accuracy (UA), overall accuracy (OA) and kappa coefficients of the MLC, SVM and proposed classifications for datasets A and B. Veg: vegetation, Shd: shadow, Bld: building, and Oth:other.

| Datasets | Classification | UA | | | | OA | Kappa |
|-----------|----------------|------|------|------|------|------|-------|
| | | Veg | Shd | Bld | Oth | | |
| Dataset A | MLC | 0.68 | 0.90 | 0.65 | 0.80 | 0.77 | 0.66 |
| | SVM | 0.69 | 0.87 | 0.73 | 0.80 | 0.78 | 0.67 |
| | Proposed | 0.80 | 0.93 | 0.74 | 0.85 | 0.83 | 0.76 |
| Dataset B | MLC | 0.68 | 0.69 | 0.66 | 0.77 | 0.71 | 0.58 |
| | SVM | 0.66 | 0.68 | 0.63 | 0.72 | 0.68 | 0.54 |
| | Proposed | 0.78 | 0.77 | 0.66 | 0.84 | 0.76 | 0.66 |

Table 2.4: Confusion matrix of the proposed classification for dataset A. PA refers to Producer Accuracy.

| Classified | Reference | | | | UA |
|------------|-----------|-------|-------|--------|------|
| | Veg | Shd | Bld | Oth | |
| Veg | 63824 | 900 | 1217 | 14313 | 0.80 |
| Shd | 231 | 53731 | 2796 | 1268 | 0.93 |
| Bld | 1219 | 2085 | 43577 | 12253 | 0.74 |
| Oth | 9829 | 2328 | 13914 | 150618 | 0.85 |
| PA | 0.85 | 0.91 | 0.71 | 0.84 | |

Table 2.5: Confusion matrix of the proposed classification for dataset B. PA refers to Producer Accuracy.

| Classified | Reference | | | | UA |
|------------|-----------|-------|-------|--------|------|
| | Veg | Shd | Bld | Oth | |
| Veg | 54678 | 339 | 1892 | 12978 | 0.78 |
| Shd | 205 | 26875 | 4920 | 2992 | 0.77 |
| Bld | 1883 | 7328 | 78348 | 31037 | 0.66 |
| Oth | 11323 | 1772 | 7516 | 111882 | 0.84 |
| PA | 0.80 | 0.74 | 0.85 | 0.70 | |

dataset A, but a slightly lower accuracy on dataset B. From the perspective of pixel-based assessment, although the proposed method yielded the highest UA for building class, it did not give a significant advantage compared with MLC and SVM. Furthermore, from Tables 2.3 and 2.4, we noticed that many shadow pixels, equal to 2796 and 4920 of datasets A and B respectively, were misclassified into building class. A total of 9829 and 11323 other pixels of datasets A and B were misclassified into vegetation class, respectively.

We assessed the results of the MLC, SVM and proposed classifications based on the modified over-classification and under-classification error indices. The results using three global error (i.e. GOC, GUC and GTC) indices are given in Tables 2.6 and 2.7, corresponding to datasets A and B, respectively. These two tables show that, for building and other classes, the proposed method produced the smallest GOC, GUC and GTC errors; for vegetation and shadow classes, it also yielded the smallest GOC and GTC errors, but slight higher errors in GUC. These two tables also show that, for both datasets A and B, MLC and SVM yielded similar GOC and GUC and GTC errors for all four classes, except a relatively large difference in the GUC error for building class on dataset B, corresponding to 0.28 and 0.45

2. Object-based urban land cover classification from VHR images

Table 2.6: Geometric accuracies of the MLC, SVM and proposed (labelled as Pro) classifications for dataset A, evaluated by GOC, GUC and GTC global error indices, corresponding to global over-classification, under-classification and total classification errors respectively.

| | GOC | | | GUC | | | GTC | | |
|-----|------|------|------|------|------|------|------|------|------|
| | MLC | SVM | Pro | MLC | SVM | Pro | MLC | SVM | Pro |
| Veg | 0.32 | 0.31 | 0.21 | 0.17 | 0.14 | 0.16 | 0.27 | 0.25 | 0.20 |
| Shd | 0.10 | 0.13 | 0.07 | 0.06 | 0.06 | 0.14 | 0.09 | 0.11 | 0.13 |
| Bld | 0.42 | 0.34 | 0.26 | 0.57 | 0.63 | 0.30 | 0.52 | 0.51 | 0.30 |
| Oth | 0.20 | 0.20 | 0.15 | 0.28 | 0.23 | 0.18 | 0.24 | 0.22 | 0.17 |

Table 2.7: Geometric accuracies of the MLC, SVM and proposed (labelled as Pro) classifications for dataset B, evaluated by GOC, GUC and GTC global error indices, corresponding to global over-classification, under-classification and total classification errors respectively.

| | GOC | | | GUC | | | GTC | | |
|-----|------|------|------|------|------|------|------|------|------|
| | MLC | SVM | Pro | MLC | SVM | Pro | MLC | SVM | Pro |
| Veg | 0.32 | 0.34 | 0.22 | 0.21 | 0.33 | 0.32 | 0.29 | 0.36 | 0.29 |
| Shd | 0.31 | 0.32 | 0.24 | 0.26 | 0.24 | 0.29 | 0.30 | 0.30 | 0.29 |
| Bld | 0.35 | 0.37 | 0.34 | 0.28 | 0.45 | 0.18 | 0.34 | 0.43 | 0.28 |
| Oth | 0.30 | 0.28 | 0.16 | 0.65 | 0.66 | 0.52 | 0.50 | 0.52 | 0.38 |

respectively. From the perspective of global total error, for dataset A, MLC produced the largest GTC error for building and other classes, equal to 0.52 and 0.25 respectively. By contrast, the proposed method produced GTC values of 0.30 and 0.17 for these two classes. For dataset B, SVM produced the largest GTC values for building and other classes, equal to 0.43 and 0.52, whereas the proposed method yielded 0.28 and 0.38 respectively.

Figure 2.13 presents the distribution of the total error index of building objects classified by the MLC, SVM and proposed methods. The dark objects that do not overlap with reference objects are referred to as false positives. The MLC and SVM classifications, shown in Figures 2.13(a1, a2, b1 and b2) have more false positives than the proposed classifications, corresponding to Figures 2.13(a3 and b3). The reference objects that fail to be extracted in a classification map are referred to as false negatives. Figures 2.13(a1, a2, b1 and b2) also show a higher number of false negatives in MLC and SVM classifications compared with the proposed classifications, shown in Figures 2.13(a3 and b3), on both datasets A and B.

2.5 Discussion

In this study we proposed a method for classifying urban land cover from VHR satellite images. The results of image segmentation revealed that integrating a BPT with energy measures provides an effective means for segmenting VHR images. Several studies have used BPT to process satellite images (Kurtz et al., 2012; Salembier and Garrido, 2000; Valero et al., 2013). Previous research has also suggested that energy measures are applicable to evaluate image partitions (Guigues et al., 2006; Philipp-Foliguet and Guigues,

2008; Serra and Kiran, 2013). So far, however, very little effort has been made to synthesize the properties of both BPT and energy measures. Therefore, the synthesis between BPT and energy measures is a key contribution of this work. The results of building extraction also confirmed that the directional relationship based on fuzzy landscape supplies effective information for distinguishing building from other man-made objects.

The strength of using BPT to represent image is that it provides a hierarchical representation for an image by which image objects can be interpreted at multiple scales. The difficulty of using BPT is however to decide a meaningful representation or a segmentation at a specific scale. This research provided a way of obtaining the relatively optimal representation of an image, by means of unsupervised segmentation evaluation. We used two energy measures ED and EC to evaluate a set of possible image segmentations represented by a BPT for choosing the segmentation that has the minimum energy. This work can be extended by allowing two energy measures ED and EC of different importance. The obtained optimal segmentation may then vary, as using different weights for ED and EC can produce image segmentation at multiple scales. Such segmentations are important for particular remote sensing applications, e.g. those considering scale issues.

The creation of a BPT of an image relies on the region model and merging criterion as well as an initial partition. The adopted region merging criterion produced acceptable results in this study, but it still faces some limitations. In particular, if a region R_i is a hole of (i.e. fully surrounded by) its neighbouring region R_j , these two regions will always merge, irrespective of their attribute values. Future research has to be conducted on improving the BPT region merging criterion. Furthermore, from creating a BPT using various initial partitions, we observed that the initial partitions derived from the same initial segmentation (e.g. SF1-3 or SM1-3) did not have an evident effect on the created BPTs if they were evaluated by energy measures. The created BPTs however showed a slight difference due to the initial segmentation methods (i.e. SF vs. SM). This difference was caused by two aspects: (1) SF1-3 initial partitions had larger ED values but smaller EC values than those of SM1-3, and (2) the ED and EC energies were considered of the same importance. As a result, the data fitting strategy (i.e. MIN2) produced the optimal segmentations at a coarser level of image detail than the straightforward method (i.e. MIN1).

The directional relationship between buildings and shadows represented by a fuzzy landscape provides effective information for building extraction. In practice, the use of shadows still faces some limitations. The shadows of buildings may not always be detectable. For example, the shadows may be casted onto buildings themselves if they are parallel to the illumination direction or onto tall buildings that stand close to each other. The shadow direction may not be specified at polar and equatorial areas. In these cases the shadows cannot be detected and hence the building cannot be detected. On the other hand, the shadows casted by trees and bridges also influence the accuracy of building extraction.

From the results of performing pixel-based and object-based accuracy assessment, the proposed method generally performed better than MLC and

SVM. Moreover, the modified over-classification and under-classification error indices aimed at dealing with one-to-many instances when matching the classified and reference objects. They provide a useful method for object-based accuracy assessment. The classification results showed that several building objects failed to be extracted by the proposed classification (Figure 2.12). Typically, for dataset A, small shadow objects, which were characterized with spectral features similar to some dark man-made objects, were not selected into the training dataset for the preliminary SVM classification. If we select these small objects into training dataset, the classification result would result in several misclassifications between shadow and dark man-made objects. To face this problem, we made a trade-off for training dataset collection. For the pixel-based assessment, we observed that for the building class, the proposed method has a similar accuracy as MLC and SVM in terms of overall accuracy and kappa coefficient. Second, the confusion matrices of the proposed classifications show that many vegetation and shadow pixels were misclassified into building and other classes, respectively. In this work, the reference data were produced by digitalization of datasets. Uncertainty involved in the reference data may contribute to problematic confusion matrices. Some studies about these issues can be found in literature (Foody, 2010; Stein et al., 2009).

2.6 Conclusions

In this study we proposed an object-based method for classifying urban land cover using Pléiades images. We combined a Binary Partition Tree (BPT) with energy minimization for image segmentation, and applied fuzzy sets to create a fuzzy landscape for building extraction. BPT provided a useful means for obtaining an optimal image segmentation without using reference data. This segmentation is based upon a hierarchical representation of an image by a BPT and an unsupervised evaluation of the image partitions by energy minimization. Modelling the directional relationship between building and shadow objects provided effective supplementary information for distinguishing buildings from other man-made objects. By constructing a fuzzy landscape, the crisp rule that a building object is adjacent to a shadow object was quantified as membership values into a fuzzy set. The membership values corresponded to the degree of satisfying the directional relationship that a building object casted a shadow object in a specific direction. Comparing with Maximum Likelihood Classification and Support Vector Machine classification, the proposed classification produced the highest overall accuracies and kappa coefficients, and the smallest over-classification and under-classification geometric errors for two datasets. The results allow us to conclude that the proposed method offered a potential solution for urban applications in need of extracting urban land cover.

2.7 Appendix

Table 2A.1: Spectral features of image objects.

| Object features | Feature description |
|-----------------|--|
| Mean | \bar{s}_i , the intensity of image object in spectral band i (i.e. Blue, Green, Red and NIR). |
| Brightness | $\frac{1}{m_s} \times \sum_{i=1}^{m_s} s_i$, m_s is the number of spectral bands. |
| Max_Diff | $\frac{\max_{i,j \in m_s} s_i - s_j }{\bar{s}}$, \bar{s} is the mean intensity of image object in all spectral band. |
| NDVI | $(\bar{s}_i - \bar{s}_j) / (\bar{s}_i + \bar{s}_j)$, i is the NIR band, j is the Red band. |

Table 2A.2: Geometrical features of image objects.

| Object features | Feature description |
|-----------------|---|
| Length/With | len/wd , measures the elongation of a polygon. |
| Border Index | $p/(2 \times (len + wd))$, describes how jagged a polygon is, the more jagged, the higher its border index. |
| Compactness | $\sqrt{4 \times a/\pi}/p$, describes how compact a polygon is, the more compact, the higher its compactness. |
| Shape Index | $b/(4 \times \sqrt{a})$, describes how smooth a polygon is, the smoother, the lower its shape index. |
| Roundness | $4 \times a/(\pi \times len^2)$, describes how similar a polygon is to an ellipse, the roundness value of a circle is 1. |
| Rectangular fit | $a/(len \times wd)$, describes how well a polygon is fitted by a rectangle, the rectangular fit value of a rectangle is 1. |
| Solidity | a/a_{con} , compares the area of the polygon to the area of its convex hull. |
| Convexity | p/p_{con} , describes how well a polygon is to be convex. |

len and wd : major and minor axes of the polygon, which are derived from an oriented bounding box containing this polygon.
 a : area of the polygon.
 a_{con} : area of the convex hull of the polygon.
 p : border length of the polygon.
 p_{con} : border length of the convex hull of the polygon.

Table 2A.3: BPT generation procedure.

| |
|---|
| Initialize a partition with M regions. |
| $t = 1$ |
| while $t < M$ do |
| Generate the Region Adjacency Graph (RAG) for defining the neighbouring relations between regions. |
| Calculate the dissimilarities of the pairs of neighbouring regions, and choose the most similar pair of neighbouring regions R_i and R_j . |
| Merge the region R_i and R_j to create a new region R_{M+t} (i.e. a node in a BPT). |
| Update the RAG by removing the edge connected by R_i and R_j , and creating the new edges connected with R_{M+t} . |
| $t = t + 1$. |
| End. |

Region-based Urban Road Extraction from VHR Satellite Images Using Binary Partition Tree

3

This chapter is based on the published paper: Li, M., Stein, A., Bijker, W., Zhan, Q., 2016. Region-based Urban Road Extraction from VHR Satellite Images Using Binary Partition Tree. *International Journal of Applied Earth Observation and Geoinformation* 44, 217–225.

Abstract

This paper provides a hierarchical method for urban road extraction. It consists of 1) obtaining the road region of interest from a VHR image, 2) hierarchically representing this road region of interest in a Binary Partition Tree (BPT), and extracting the roads based on this BPT at hierarchical levels. Besides using two existing geometrical features (i.e. compactness and elongation), we define two other structural features based on orientation histograms and morphological profiles to guide the region merging of BPT. The morphological profiles are constructed using a series of path openings, which facilitate modeling linear or curved structures. The proposed method was applied to two types of VHR images with different urban settings, corresponding to a Pléiades-B image of Wuhan, China, and a Quickbird image of Enschede, the Netherlands. Experimental results showed that the proposed method was able to group adjacent small segments that have high spectral heterogeneity and low road-like geometrical properties to form more meaningful roads sections, and performed superior to the existing methods. Furthermore, we compared the proposed method with two other existing methods in the literature. We conclude that the proposed method can provide an effective means for extracting roads over densely populated urban areas from VHR satellite images.

Keywords: Binary Partition Tree, fuzzy aggregation, morphological profiles, road extraction, very high resolution images.

3.1 Introduction

A subject of considerable interest for information extraction from remote sensing images is the extraction of roads. It is essential for many applications, such as transportation management, geodata updating, and land cover and land use analysis. Road extraction from remote sensing images can be traced back to Bajcsy and Tavakoli (1976), and has been applied to a large variety of sensors covering a wide range of spatial resolutions (Mena, 2003). Very High Resolution (VHR) satellite images have a spatial resolution below 1 m, from which roads exhibit as image regions with more detailed spatial information compared with low or mid-low resolution satellite images. In particular, the growing availability of VHR images, e.g. QuickBird, GeoEye, WorldView and Pléiades images, has caused an increase in the number of studies on road extraction from VHR images. Nonetheless, these studies fail to obtain an ideal result, particularly in complex urban areas where road obstacles, e.g. from vehicles, trees and shadows are commonly found, thus affecting the accuracy of road extraction.

Road extraction methods are found regularly in remote sensing literature. According to Mena (2003), road extraction methods are categorized as road tracking (Vosselman and Knecht, 1997; Hu et al., 2007; Movaghati et al., 2010; Tang et al., 2014), active contour models (snakes) (Ravanbakhsh et al., 2008; Butenuth and Heipke, 2012), mathematical morphology (Zhang et al., 1999; Soille and Pesaresi, 2002; Valero et al., 2010), segmentation and classification (Song and Civco, 2004; Bouziani et al., 2010; Shi et al., 2014), and knowledge-based methods (Baltsavias, 2004). Road extraction from images, similar to e.g. blood vessel detection from medical images, is also a well-known issue in computer vision (Chai et al., 2013). Chai et al. (2013) distinguished three types of approach: pixel-based (Stoica et al., 2004; Mnih and Hinton, 2010; Montoya-Zegarra et al., 2014), line-based (Lacoste et al., 2005), and graph-based (Gerke et al., 2004; Turetken et al., 2012; Ünsalan and Sirmacek, 2012; Turetken et al., 2013; Wegner et al., 2013). A road extraction system often involves several different types of techniques. There is no ideal road extraction for all types of land structures, in rural, suburban and urban areas. Besides using the spectral information of roads, many studies have stressed the importance of using the geometrical, contextual, and structural knowledge for road extraction (Baltsavias, 2004).

In a VHR image, roads appear as image regions, and hence a region-based approach could be of interest. For example, Grote et al. (2012) presented a region-based method for suburban road extraction using high resolution color infrared images as well as a digital surface model (DSM). The authors employed a set of radiometric and geometric features to group small road regions to form meaningful road-like regions. This method, however, involves a rather large number of parameters and was not tested on just a single VHR satellite image, but also involved availability of a DSM.

This paper aims to provide a novel region-based method for urban road extraction from VHR images. The proposed method uses a Binary Partition Tree (BPT) (Salembier and Garrido, 2000; Valero et al., 2013) to hierarchically represent road regions, and relies on the road region of interest

3. Urban road extraction from VHR images by a BPT

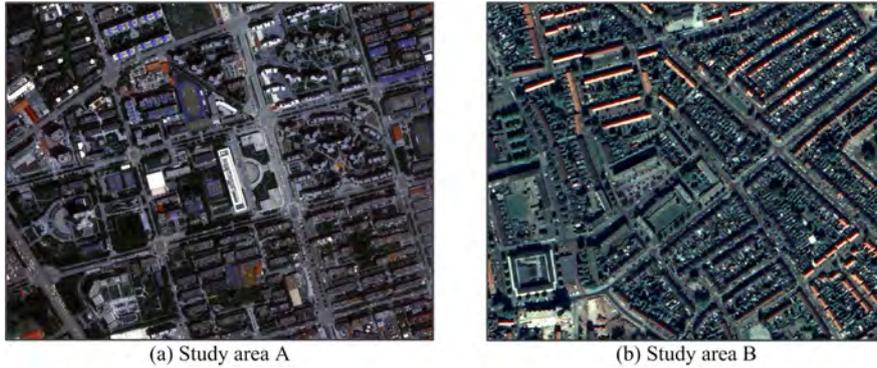


Figure 3.1: (a) The Pléiades image (with true color composite) of Wuhan, China (study area A). (b) The Quickbird image (with true color composite) of Enschede, the Netherlands (study area B).

(RoadROI) preliminarily obtained from an image. The RoadROI highlights the road information, whereas it reduces the effects of vegetation, shadow and buildings on roads. In particular, by eliminating buildings via shadows, the risk of confusing between buildings and roads because of similar spectral characteristics is reduced. The construction of a BPT uses two widely used geometrical features, i.e. compactness and elongation indices, and two newly defined structural features based on orientation histogram and morphological profiles. An automatic strategy to obtain the RoadROI from an image is provided.

The remainder of this paper is organized as follows. Section 3.2 and Section 3.3 describes the study area and used data, and illustrates the proposed method for road extraction, respectively. Section 3.4 presents experimental results and related analysis, followed by discussion in Section 3.5 and the conclusions of this research in Section 3.6.

3.2 Study area and data

The proposed method was investigated using two different VHR images over urban areas, namely study areas in Wuhan and Enschede (Figure 3.1). For study area A, the road extraction is part of a large study on land use changes of Wuhan, China, being a dynamic area. A piece of a Pléiades image was acquired on 11 July 2013. This image was acquired with an azimuth angle of 140.8° , four multispectral bands of 2 m spatial resolution, a panchromatic band of 0.5 m, and has a size of 1599×2019 pixels. This 0.5 m panchromatic band was resampled from the 0.7 m raw panchromatic band by the image provider using spline resampling. The multispectral bands were fused with the panchromatic band using the Gram-Schmidt pan-sharpening method (as same as study area B). Study area B is a more stable and established urban area. We used a Quickbird image acquired on 21 September 2006. This image has an azimuth angle of 172° , four multispectral bands of 2.4 m spatial

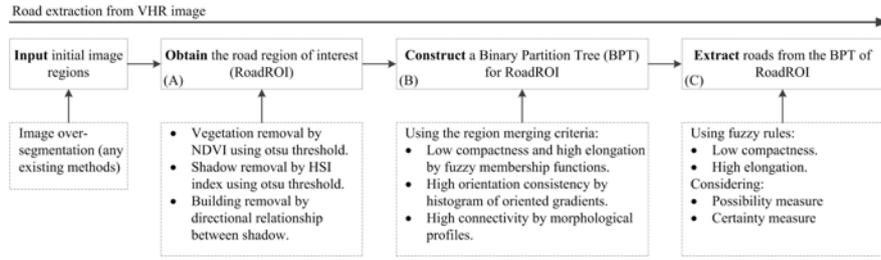


Figure 3.2: Workflow of the proposed road extraction method.

resolution, a panchromatic band of 0.6 m, and a size of 928×1089 pixels. These two study areas show that roads have a similar spectral reflectance as buildings, parking lots, squares and other built-up areas, making road extraction intrinsically difficult. Furthermore, the vehicles on roads and the trees surrounding roads give rise to obstacles. Both of these two study areas therefore make the road extraction from VHR images challenging. The reference road map of study area A was produced based on the official land use map by manually digitizing a part of small roads. The reference map of study area B was produced from the public Open Street Map.

3.3 Methods

3.3.1 Obtain the road region of interest

The workflow of the proposed road extraction is presented in Figure 3.2. It is a region-based method, which starts from an initial segmentation, i.e. an over-segmentation case where a single object is partitioned into several subdivisions. Vegetation is removed first, followed by building shadows and buildings. In this way, the road region of interest (RoadROI) is obtained. This RoadROI is used to construct a BPT for hierarchically representing road-like regions. In doing so, we make use of geometrical and structural information that is used for extracting roads based on fuzzy logic.

Let I be an image, and $P = \{R_i\}$ be an initial segmentation with $i = 1, 2, \dots, M$ regions R_i . Initial segmentation is obtained by existing segmentation methods. Here we use the Fractal Net Evolution Approach (FNEA) (Benz et al., 2004) implemented in eCognition Software by giving a small scale parameter. For each region R_i , we first smooth the within-pixel values by using the mean value of R_i with respect to a spectral band. In other words, each pixel in region R_i is replaced by the mean of the pixel values in that region. The smoothing process aims to reduce image noise with respect to vegetation, shadow, and building regions obtained by pixel-based methods. The smoothed image is therefore used for obtaining the RoadROI. We remove the vegetation regions based on the Normalized Difference Vegetation Index (NDVI) using Otsu's threshold method (Otsu, 1979). Next, shadow regions are automatically detected by using the method proposed by Teke et al. (2011), where a Hue-Saturation-Intensity (HSI) color space

is converted from a Near-Infrared-Red-Green false color space to formulate an index $(Saturation - Intensity)/(Saturation + Intensity)$, by which the Otsu's threshold is applied for shadow detection. Dark objects, however, might be mis-detected as shadows if we directly use this method for the entire image. To improve the detection performance, we subset the images into small patches for improving the bimodality of the image histogram for Otsu thresholding (Adeline et al., 2013).

Moreover, as buildings have similar spectral and textural properties as roads, we make use of the directional relationship between buildings and their shadows to exclude the buildings from the image. The directional relationship is modeled by a fuzzy landscape of building shadows. The detailed description and computation of a fuzzy landscape can be found in Li et al. (2015a). A reference (shadow) region R_{ref} and a direction with angle β is considered. Then for each image pixel $x \in I$, the fuzzy landscape $H_\beta(R_{ref})(x)$ specifies a membership function to the reference region R_{ref} . The $H_\beta(R_{ref})(x)$ ranges from 0 to 1, and specifies the degree of satisfying the directional relationship of $x \in I$ to R_{ref} . This fuzzy landscape is further smoothed at the region-level by averaging the within pixel values of an image region R_i . Based upon the smoothed fuzzy landscape H , the building regions are automatically detected if their directional relationship is larger than 0.5. By using this method, other shadows, i.e. mainly tree shadows, are removed from the set of that can then be used for building detection. Removal of tree shadows is further handled by using directional relationships between trees (i.e. vegetation) and their shadows. To do so, we first obtain the vegetation boundaries adjacent to shadows, then construct another fuzzy landscape that has the opposite direction angle $180^\circ - \beta$ where the pixel value specifies the degree that a shadow pixel is casted by a tree. Similarly, tree shadow regions can be detected if their directional relationship is larger than 0.5. The rest of image regions that are not vegetation, buildings and their shadows will be used for RoadROI BPT construction.

3.3.2 Construction of a BPT for RoadROI

A BPT provides a tree-structure for a region-based representation of an image, and has been used for image processing, segmentation and objects extraction (Salembier and Garrido, 2000; Sigurdsson et al., 2014; Valero et al., 2015). Let T be a BPT, and v be a node of T , corresponding to an initial image region. As shown in Figure 3.3, the top node R11 is the root of T with two children nodes R10 and R6. The node R10 is referred as a branch node and the R6 is a leaf node as it has no children nodes.

3.3.2.1 Region model

The construction of a BPT involves two key components: the region model and region merging. The region model specifies how regions are represented and how to compute the union of two adjacent regions (Kurtz et al., 2012). Two shape features, namely compactness and elongation indices, and two newly defined structural features, namely the orientation histogram and

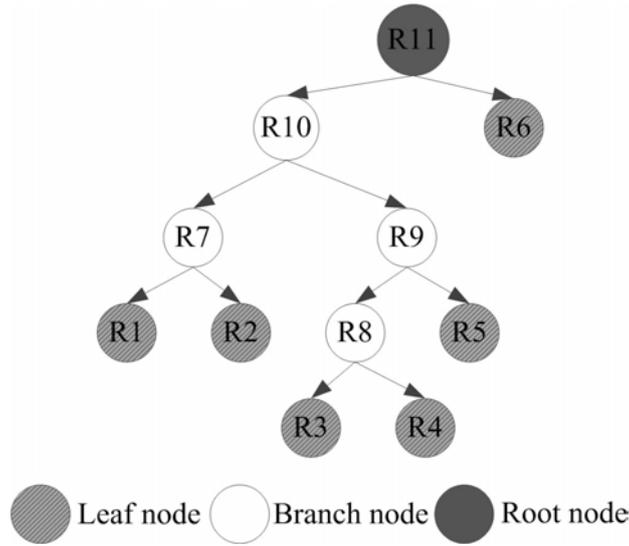


Figure 3.3: Tree-structure representation of a BPT sample with six initial regions.

connectivity, are used to define the region model with respect to an image region in RoadROI. The compactness cm and elongation eln indices (Bouziani et al., 2010) are defined for each region R_i as

$$cm_i = 2 \cdot \sqrt{\pi \cdot a_i} / p_i \quad (3.1)$$

$$eln_i = a_i / (len_i^2) \quad (3.2)$$

where a_i , p_i and len_i are the area, perimeter and length (i.e. the major axis of a polygon) of R_i , respectively.

Inspired by the Histogram of Oriented Gradients (HoG) (Dalal and Triggs, 2005), for a region R_i , we calculate an orientation histogram with k_b bins, labeled HoG_i , to represent the orientation distribution of R_i . Differing from the original HoG in Dalal and Triggs (2005), for a region R_i , the local neighborhood of calculating its HoG_i is determined by the size of the bounding box of R_i and a given cell size k_c . Specifically, the bounding box of R_i is buffered if either the width or the length of this bounding box is less than k_c . By doing so, the minimum local neighborhood of a HoG_i is guaranteed. The rest of the calculation of a HoG_i is the same as the original HoG calculation. The reader is referred to Dalal and Triggs (2005) for more details about HoG calculation.

Morphological profiles (MP) have been widely used to model spatial information of the images (Ghamisi et al., 2015). An MP is created by applying a series of traditional openings and closings with a structuring element (SE) of increasing size to an image. A major shortcoming of using traditional openings or closings for building an MP, however, is the difficulty in

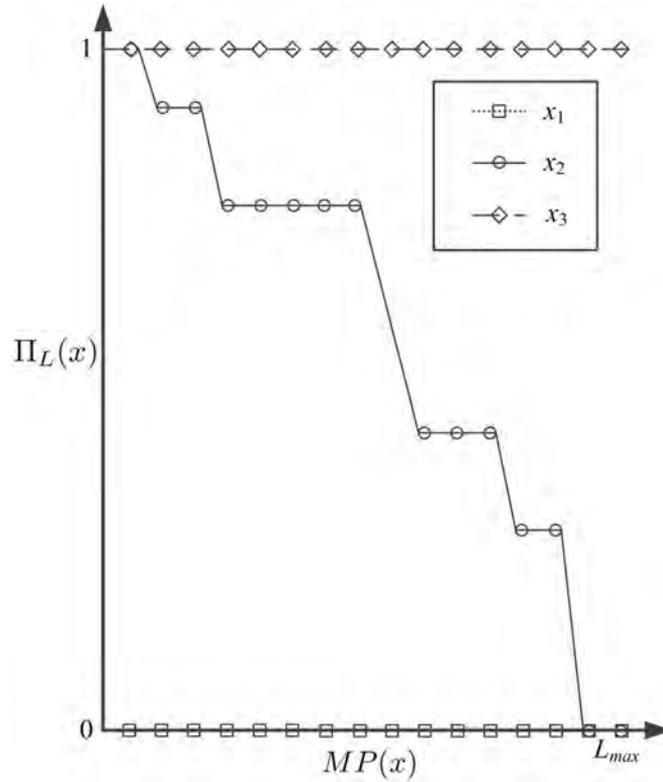


Figure 3.4: Three MPs of pixel x_1 , x_2 , x_3 created by using path openings with the length L ranging from 0 to L_{max} pixels. Specifically, x_1 , x_2 , x_3 refer to not a road pixel, probably a road pixel, and a road pixel in a RoadROI.

determining a proper SE for conducting the opening or closing operation. Furthermore, traditional openings and closings fail to extract curved structures, for instance, a curved road in an image. To overcome this shortcoming, Heijmans et al. (2005) proposed two variants, called path openings and closings, to facilitate extracting pixel structures that are locally oriented but not perfectly straight, from binary and gray-level images. It removes the pixels that are not connected in an adjacency graph under a given length and orientation. The theoretical foundations of path openings and closings are detailed given in Heijmans et al. (2005). The L -tuple (x_1, x_2, \dots, x_L) is called a δ -path of length L if $x_k \mapsto x_{k+1}$, or equivalently, if $x_{k+1} \in \delta(\{x_k\})$, for $k = 1, 2, \dots, L - 1$. The \mapsto refers to a directed adjacency relation, meaning that there is an edge going from x_k to x_{k+1} . Let the set of all δ -paths of length L be Π_L , and let the set of δ -paths of length L contained in a subset X of an image I be $\Pi_L(X)$, $X \subset I$. The operator $\alpha_L(X)$ called the path opening is defined as the union of all δ -paths of length L contained in X . We apply a series of path openings with a length of increasing size to form an MP on the pixels of the RoadROI. The morphological profile $MP(x)$ of a

pixel x is defined as

$$MP(x) = \{\Pi_L(x)\}, L \in \{1, \dots, L_{max}\}, \quad (3.3)$$

where L_{max} indicates the longest length of the used path openings. We exemplify the characteristics of the MPs with respect to different types of pixels, namely non-road, possible road and road pixels (Fig 3.4).

Based upon the MP of RoadROI, we define a connectivity feature to measure to which degree an image region is linear or curved. The connectivity feature con_i of R_i is defined as

$$con_i = \frac{1}{n_i} \sum_{p=1}^{n_i} \sum_{L=0}^{L_{max}} w_L \cdot \Pi_L(x_p), x_p \in R_i, \quad (3.4)$$

where w_L refers to the weight of a path opening with length L being considered in con_i . For a simple linear case, the w_L equals to $\frac{1}{L_{max}}$. In addition, weight values calculated from a Gaussian distribution can also be used for w_L .

3.3.2.2 Region merging

The region merging criterion determine the order of pair-wise regions being merged in a BPT creation. The two geometrical features of the union of adjacent regions are quantified via a Z-shape (or spline-based) fuzzy membership function,

$$f(x; a, b) = \begin{cases} 1, & x \leq a \\ 1 - 2 \cdot \left(\frac{x-a}{b-a}\right)^2, & a \leq x \leq \frac{a+b}{2} \\ 2 \cdot \left(\frac{x-a}{b-a}\right)^2, & \frac{a+b}{2} \leq x \leq b \\ 0, & x \geq b \end{cases} \quad (3.5)$$

$$\mu_{1,ij} = f(cm_{ij}, a_1, b_1), \quad (3.6)$$

$$\mu_{2,ij} = f(eln_{ij}, a_2, b_2), \quad (3.7)$$

where $\mu_{1,ij}$ and $\mu_{2,ij}$ refer to the membership values of cm_{ij} or eln_{ij} respectively. Theoretically, the values of cm_{ij} and eln_{ij} have the range of $(0, 1)$. Here the variable a_1 and a_2 are equal to 0 for both cases.

Next, we use histogram intersection to quantify the orientation similarity between two adjacent regions. Let HoG_i , HoG_j and HoG_{ij} be the orientation histograms of the objects R_i , R_j , and their union R_{ij} , respectively. The orientation similarity $\mu_{3,ij}$ of two adjacent regions is defined as,

$$\mu_{3,ij} = \max(f_{bin_i}(R_i) - f_{bin_i}(R_j), f_{bin_j}(R_j) - f_{bin_j}(R_i)) \quad (3.8)$$

where $f_{bin_i}(R_i)$ (resp. $f_{bin_j}(R_j)$) is the maximum normalized frequency (equal to 1) of the HoG_i (resp. HoG_j). Here, we normalize a histogram by

3. Urban road extraction from VHR images by a BPT

dividing it using the largest value of this histogram. The subscript bin_i (resp. bin_j) indicates a bin location that has the maximum normalized frequency of the HoG_i (resp. HoG_j). The $f_{bin_i}(R_j)$ refers to the normalized frequency of HoG_j at the bin location bin_i . Likewise, the $f_{bin_j}(R_i)$ is the normalized frequency of the HoG_i at the bin location bin_j .

The MP is used to measure the connectivity of two adjacent regions. Let con_i and con_j be the connectivity features of R_i and R_j . The connectivity of two adjacent regions $\mu_{4,ij}$ is defined as

$$\mu_{4,ij} = (con_i \cdot a_i + con_j \cdot a_j) / (a_i + a_j) \quad (3.9)$$

where a_i and a_j are the areas of R_i and R_j , respectively.

In summary, we define region merging criteria as

$$O_{ij} = w_{rm} \cdot \sqrt{\mu_{1,ij} \cdot \mu_{2,ij}} + (1 - w_{rm}) \cdot \sqrt{\mu_{3,ij} \cdot \mu_{4,ij}} + \epsilon_{rm}, \quad (3.10)$$

$$\epsilon_{rm} = |wd_{ij} - \max(wd_i, wd_j)| / wd_{ij}, \quad (3.11)$$

where $\sqrt{\mu_{1,ij} \cdot \mu_{2,ij}}$ and $\sqrt{\mu_{3,ij} \cdot \mu_{4,ij}}$ are two terms representing the geometrical and structural features, respectively, and the parameter w_{rm} weights these two terms. In this paper, we take $w_{rm} = 0.5$. The parameter ϵ_{rm} aims at constraining the width of the two adjacent merged regions, where wd_i, wd_j and wd_{ij} are the width of the regions R_i, R_j and R_{ij} respectively. A high value of O_{ij} gives high priority to the two adjacent objects R_i and R_j to be merged. Based upon the defined region model and region merging criteria, the BPT of a RoadROI is created.

3.3.3 Road extraction

Based upon the constructed RoadROI BPT, road extraction is performed. The initial image regions are merged to form more road-like regions represented by BPT nodes. Given an image region R_v represented by a BPT node v , the membership degree of R_v being a road region is quantified by two fuzzy membership values $\mu_{1,v}$ and $\mu_{2,v}$. We then apply the method of Sebari and He (2013), to detect road regions from a BPT based on these two fuzzy rules. Furthermore, the adopted method also quantifies the uncertainty of extracted road regions. Specifically, a possibility measure $\prod(R_v)$ represents the possibility that an image region R_v belongs to road,

$$\prod(R_v) = \max(\mu_{1,v}, \mu_{2,v}) \quad (3.12)$$

$$\prod(\overline{R_v}) = \max(1 - \mu_{1,v}, 1 - \mu_{2,v}) \quad (3.13)$$

where $\prod(\overline{R_v})$ denotes the possibility that the region R_v belongs to non-road. In addition, a necessity measure $N(R_v)$ is defined to represent the certainty that an image region R_v is road,

$$N(R_v) = 1 - \prod(\overline{R_v}) \quad (3.14)$$

$$N(\overline{R_v}) = 1 - \prod(R_v) \quad (3.15)$$

where $N(\overline{R_v})$ denotes the possibility that the region R_v belongs to non-road. The image region R_v is therefore detected as a road class if $\prod(R_v) > \prod(\overline{R_v})$ and $N(R_v) > N(\overline{R_v})$.

3.3.4 Quality assessment

We compare the proposed method with two other road extraction methods found in the literature. We first use the Orfeo ToolBox (Inglada and Christophe, 2009), labeled as OTB method, to extract roads from both Pléiades and Quickbird study images. The Orfeo ToolBox is an open source C++ library for remote sensing images processing, which is developed by the French space agency to promote the use of Pléiades images. Hence, this toolbox can be well used to compare the results between the proposed and existing methods, particularly using Pléiades images. In addition, we compare our extracted roads with the method from Valero et al. (2010), labeled the S.V. method, that investigated the use of MP for road extraction. Because the proposed method also uses MP, we consider it a proper option for the comparison purpose. For implementation of the OTB and S.V. methods, we set the corresponding parameters as follows; For OTB extraction, an input pixel of roads is used for each study area, see Figure 3.6; other parameters are kept the same as suggested in the user guide of Orfeo Toolbox. For S.V. extraction, the MGL parameter is obtained by using a path opening operation of length 400 pixels; the L_{min} equals to 400 pixels; and the T parameter is equal to 50 which is the same as used in the literature (Valero et al., 2010).

We evaluate the extracted roads using three accuracy measures, namely completeness, correctness, and quality, used by Clode et al. (2007),

$$completeness = \frac{TP}{TP + FN}, \quad (3.16)$$

$$correctness = \frac{TP}{TP + FP}, \quad (3.17)$$

$$quality = \frac{TP}{TP + FP + FN}, \quad (3.18)$$

where TP indicates the true positive and is the number of pixels existing in both reference and extracted roads; FN indicates the false negative and is the number of pixels existing in the reference roads, but not in the extracted roads; FP indicates the false positives and is the number of pixels existing in the extracted roads, but not in the reference roads.

3. Urban road extraction from VHR images by a BPT

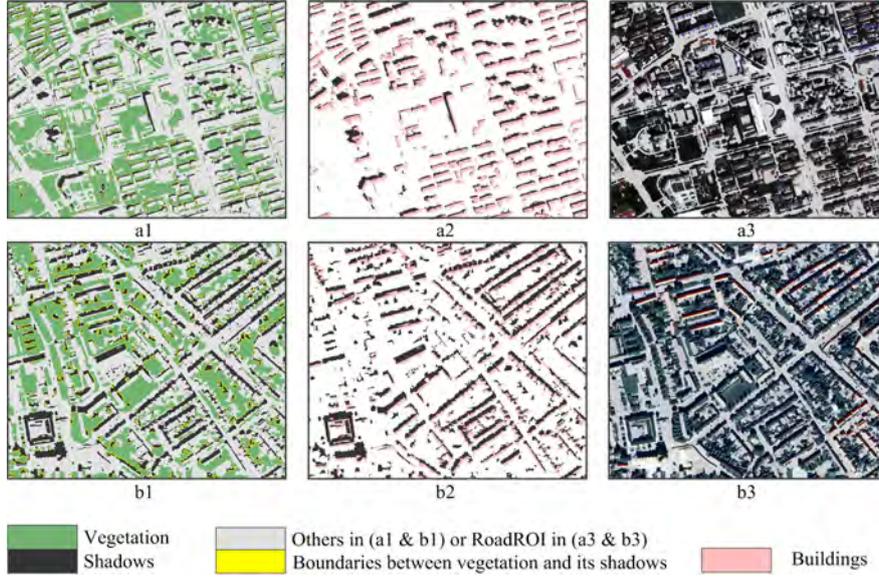


Figure 3.5: The obtained RoadROIs of both study areas A and B, corresponding to the top and bottom panels, respectively. (a1 and b1) are the detected vegetation and shadows by Otsu’s threshold. (a2 and b2) are the detected buildings by directional relationship between buildings and their shadows. (a3 and b3) are the obtained RoadROIs overlapping with the corresponding true color VHR images.

Table 3.1: Parameter settings associated with the steps of road extraction. Internal parameters refer to those described in this paper. External parameters refer to those associated with existing algorithms but were not in detail described in this paper.

| Parameters setting | Description |
|-----------------------------------|--|
| Obtain RoadROI | |
| $Scale = 40$ | (External) Scale parameter in eCognition software for creating an image segmentation. Other parameters were set as default. |
| $\omega = 140.8^\circ, 172^\circ$ | (External) Azimuth angle of a VHR image, used for calculating the direction parameter β (Li et al., 2015a). |
| Construct RoadROI BPT | |
| $b_1 = 0.6; b_2 = 0.4$ | (Internal) Parameters for defining Z-shape membership functions. |
| $k_c = 50; k_b = 18$ | (Internal) Parameters for calculating a HOG_i . |
| $L_{max} = 400$ | (Internal) Parameter for calculating a MP. |
| $w_{rm} = 0.5$ | (Internal) Parameter for defining the region merging criteria. |

3.4 Results

3.4.1 Obtain the road region of interest

Figure 3.5 shows the results of obtaining the RoadROIs at both study areas A and B. Figures 3.5a1 and 3.5b1 give the detected vegetation and shadows by Otsu’s threshold. To erase the tree shadows, we constructed a fuzzy land-

scape to model the directional relationship between shadows and their adjacent trees (i.e. vegetation in this work). Table 3.1 gives the size of the used fuzzy structuring element, as well as other parameter settings associated with the step of obtaining the RoadROI. The rest of the detected shadows, corresponding to building shadows, were used to detect buildings. Figures 3.5a2 and 3.5b2 display the detected building shadows. The RoadROIs of study areas A and B were obtained by masking the detected vegetation, building shadows, and buildings (Figures 3.5a3 and 3.5b3).

3.4.2 Construction of a BPT for RoadROI

The construction of a RoadROI BPT involves in calculating two shape features, namely compactness and elongation, as well as two structural features, namely orientation histogram and connectivity. Table 3.1 lists the parameter settings of this BPT construction. To define the parameters for the Z-shape membership functions in equations (3.6 and 3.7), we used the compactness and elongation information about roads in the literature, i.e. the compactness of most roads was lower than 0.3 and the elongation was lower than 0.2. We set the parameters, for instance, $b_1 = 0.6$ to make sure that the possibility of a region with compactness less than 0.3 is larger than 0.5. To calculate an orientation histogram of an object, we set the minimum neighborhood k_c equal to 50 pixels, and chose the number of bins k_b equal to 18. To calculate the connectivity feature, we set the maximum length L_{max} of path opening equal to 400 pixels.

3.4.3 Road extraction

Figure 3.6 gives the extracted results at study areas A and B by the proposed and the OTB and S.V. methods. The proposed method gave the best extraction as compared to OTB and S. V. for both study areas A and B. Figures 3.6a1 and 3.6b1 show that the roads extracted by OTB exhibit relatively low likelihood at both study areas A and B. For the roads with heavy obstacles like vehicles and vegetation, the OTB gave the worst extraction performance followed by S.V. for these two study areas, see yellow boxes A and C in Figures 3.6a1-3.6a3 and yellow ellipses F in Figure 3.6. The proposed method was able to merge those small regions caused by the appearance of vehicles to form more meaningful road-like regions. Figures 3.6a3 and 3.6a4 show that a large part of buildings and other built-up areas, particularly those close to roads, were extracted as roads, see yellow box B in Figure 3.6a3 and yellow ellipse E in Figure 3.6b3, because those adjacent buildings and other built-up areas had similar spectral reflectance as roads. By using the proposed method, some adjacent buildings were removed in the first step. Because built-up regions generally had a low elongation and a high compactness, these regions were less likely to be merged as more meaningful road-like segments by the proposed method. As shown in the yellow box B in Figure 3.6a2 and yellow ellipse D in Figure 3.6b2, however, several building were mis-extracted as roads. In fact, these buildings were fully surrounded by trees, and no building shadows were detected. Hence,

3. Urban road extraction from VHR images by a BPT

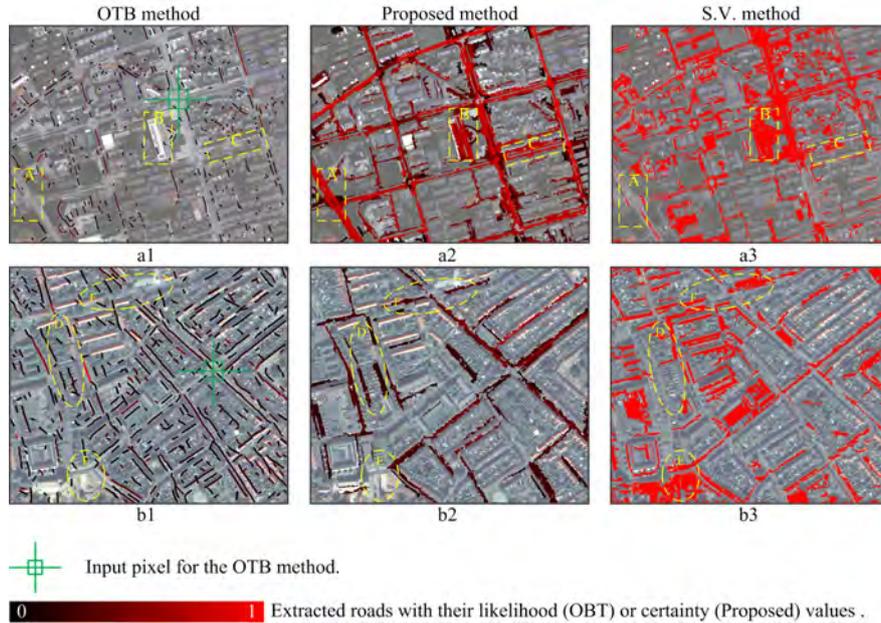


Figure 3.6: Road extraction by the proposed and other two existing methods. (a1 and b1) , (a2 and b2), and (a3 and b3) are the results produced from the OTB, proposed, and S.V. methods. The proposed and OTB methods also provide the extracted results with associated uncertainty measures, corresponding to the certainty and likelihood values respectively.

these buildings failed to be detected by shadow information at the stage of obtaining the RoadROI. In addition, these buildings were similar to roads, which have a high elongation and low compactness. In addition to road extractions at study areas A and B, the reader is referred to the supplementary data for road extraction on a large VHR image (Figure 3A.1. in Appendix

3.4.4 Quality assessment

Figure 3.7 shows the evaluation of the extracted roads of the proposed method as compared with the S.V. method. We however did not evaluate the extracted roads by OTB method, because it extracts the roads as lines instead of regions. Further, as shown in Figure 3.6, OTB produced the worst results from a visual inspection.

Figure 3.7 shows that road extraction at study area A gave a better performance than at study area B. This figure also shows that the proposed method performed better than the S.V. method at both study areas A and B. For some regions with evident obstacles like vehicles and vegetation, see the yellow boxes A and C in Figures 3.7a1-3.7a3, the S.V. method yielded a high proportion of false negatives. The reason was that, for instance, the vehicles on roads increased the spectral heterogeneity which affected the rep-

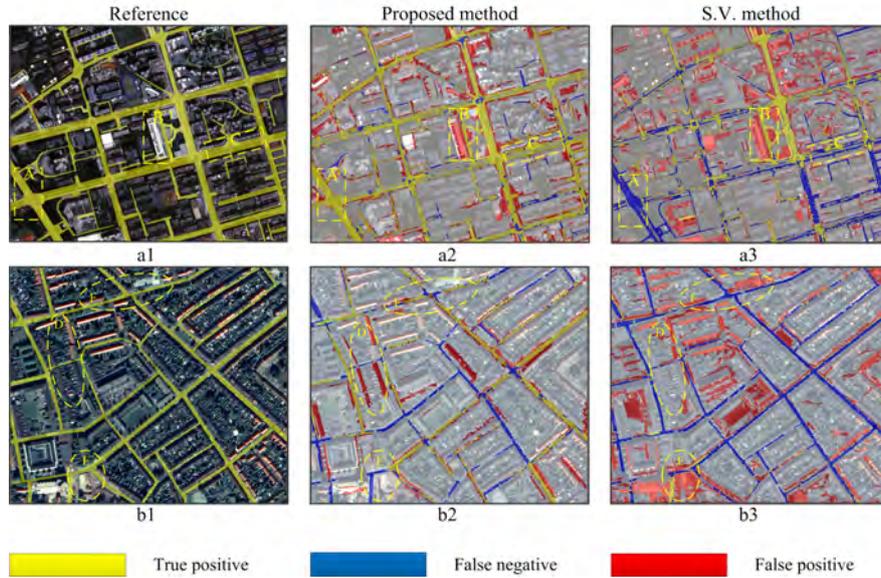


Figure 3.7: Evaluation of the extracted roads, and compared with the S.V. method. (a1 and b1) are the reference maps at study areas A and B. (a2 and b2) and (a3 and b3) refer to the valuation of extracted results by the proposed and S.V. methods, in terms of true positive, false negative and false positive measures.

resentation of linear structures by path openings. In contrast, the proposed method was able to merge small regions to form larger, more meaningful regions so that the linear structures of roads were retained. For some regions adjacent to buildings and other built-up regions, see the yellow ellipse E in Figures 3.7b1-3.7b3, the S.V. method yielded a higher number of false positives as compared with the proposed method. The reason was that these areas had a similar spectral reflectance as roads. It is thus difficult to separate roads from their surroundings just relying on spectral characteristics. The reason was that during the first step of the proposed method, buildings were removed to reduce the effects. In addition, those built-up areas were initially segmented to regions with low elongation and high compactness properties which were constrained in the region merging process of a BPT creation. Nonetheless, some buildings were mis-extracted as roads, causing false positives in the evaluation, see yellow box B in Figure 3.7a2 and yellow ellipse D in Figure 3.7b2. The reason was that these buildings were not removed from the corresponding RoadROI, and had similar geometrical properties as roads, i.e. high elongation and low compactness. Furthermore, the final road extraction only considered these two geometrical properties.

Table 3.2 gives the quantitative evaluation of the roads extracted by the proposed and S. V. methods, in terms of completeness, correctness, and quality measures. This table denotes that the roads extracted by the proposed method had a higher accuracy than the S. V. method for both study areas

3. Urban road extraction from VHR images by a BPT

Table 3.2: Evaluation of the extracted roads by the proposed and S.V methods using completeness, correctness, and quality measures. The A and B refer to the study areas A and B respectively.

| | Proposed method | | S.V. method | |
|--------------|-----------------|--------|-------------|--------|
| | A | B | A | B |
| Completeness | 0.8592 | 0.5766 | 0.4799 | 0.1283 |
| Correctness | 0.5875 | 0.4690 | 0.3758 | 0.0695 |
| Quality | 0.5359 | 0.3489 | 0.2670 | 0.0472 |

A and B. The extracted roads by either the proposed or the S. V. method at study areas B had a lower accuracy than those at study area A, in terms of these three accuracy measures. Regarding these three accuracy measures, the completeness of the extracted roads was higher than correctness and quality measures. For study area B, although the proposed method produced a better extraction than the S. V. method, the extracted roads were still evaluated with low accuracy, which might be caused by the following reasons: (1) this study area covered a dense urban area, where the buildings were arranged with small spaces between them, and (2) a large part of small roads were fully blocked by vegetation and shadows, and the image regions of these small roads were not connected and could not be grouped to more meaningful roads.

3.5 Discussion

This study investigated the use of BPT for region-based urban road extraction from VHR images. Our results confirm that BPT provides an effective representation for urban roads in VHR images, to obtain a set of road-like regions, which have high elongation and low compactness, for road extraction. Moreover, the BPT constructed in this work is application driven. This differs from a more common data driven BPT construction. To the best of our knowledge, it is hard to construct an effective BPT for various applications. An application driven BPT is more suited for a specific extraction task like road extraction, because it uses a specific region model and a merging criterion. In addition, an application driven BPT can make use of expert's knowledge for BPT construction. Therefore, this paper uses fuzzy logic to model expert's knowledge. Many studies on road extraction from remote sensing images can be found in the literature, particularly the recent work by Grote et al. (2012), who also investigated region-based road extraction by grouping image segments into more meaningful road sections. The difference exists in several aspects, e.g. (1) the proposed method starts from an over-segmentation, which can be done by any existing segmentation methods, (2) DSM data is not used, and (3) less parameters are used.

From the examples in this study it is clear that buildings have an effect

on road extraction, in particular those which are adjacent to roads or have similar shapes as roads. This finding was also confirmed by previous work (Mena, 2003; Grote et al., 2012). In this study, the proposed road extraction relies on a RoadROI, corresponding to image regions excluding vegetation, buildings and building shadows. The RoadROI places the emphasis on road information, whereas it first reduces the effect of vegetation, shadow and buildings on roads; it is obtained in an automated way. The definition of the RoadROI, however, could be dependent on study images. If an image shows many building shadows on roads, as in this study, we suggest the corresponding RoadROI to include building shadows, whereas otherwise, we suggest it to exclude building shadows. Future work could focus on analyzing the effect of building shadows on extracted roads, and could provide quantitative information to end-users whether a RoadROI should include building shadows or not. Moreover, because tree shadows affect building detection by means of shadow information (Li et al., 2015a), they are removed from the set of detected shadows that can be used for building detection.

A Z-shape fuzzy membership function is used to quantify to which degree a pair of adjacent regions should be merged according to a geometrical feature. Other fuzzy membership functions (e.g. linear or Gaussian functions) (Zadeh, 1994) can also be used for this quantification. Experimentally, there is no evident difference between different membership functions in this task. The question is to assign proper parameters for the chosen membership function. We used the geometrical information of roads in the literature (Bouziani et al., 2010), i.e. the compactness of most roads was lower than 0.3 and the elongation was lower than 0.2, corresponding to $b_1 = 0.6$ and $b_2 = 0.4$. Our results again confirm this knowledge. We also defined two new structural features, namely the orientation histograms and connectivity. The computation of the orientation histogram HOG_i of region R_i needs to specify the size k_c of the minimum local neighborhood and the number k_b of bins, where the k_c mainly guarantees the minimum local neighborhood for the orientation histograms of small regions. The effect of k_c and k_b on the computation of orientation histograms can be found in the literature (Dalal and Triggs, 2005). The computation of the connectivity con_i of region R_i needs to specify the longest length L_{max} of the used path openings. A smaller value of L_{max} may lead to a non-road region with a relatively higher value of connectivity, while a larger value of L_{max} may give a small-road region with a relatively lower value of connectivity. The region merging criterion in equation 3.10 involve the combination of the geometrical and structural terms, where $w_{rm} = 0.5$. A larger w_{rm} gives more importance on the geometrical term, and vice versa. In the literature, other alternative methods can be used for this combination, for example, the decision fusion method by Fauvel et al. (2006). Our method, however, can also achieve comparable results by a relatively straightforward combination.

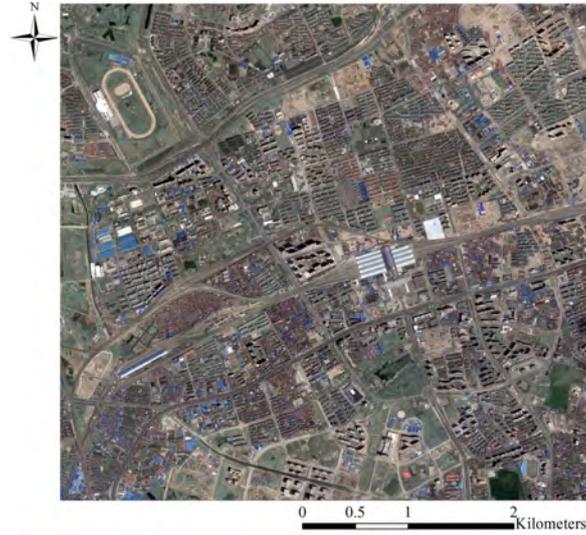
Road regions are hierarchically represented by a BPT, from which road regions are assigned as roads or non-roads using two fuzzy rules, referring to elongation and compactness properties. Although our results gave the best extraction compared with two other existing methods, more additional rules may improve the extraction performance. Moreover, new rules can also be

easily integrated in the decision step, by means of possibility and necessity measures. The results of study area B showed that, for those small roads fully blocked by vegetation and building shadows, the corresponding image regions failed to be merged. Because, no adjacency relationships between two blocked regions were defined in the Region Adjacency Graph (RAG). Future improvements can be conducted to deal with this problem, for example, if the distance of two blocked regions is smaller than a specific value, then these two regions can be seen as adjacent. In this paper, we compared the proposed method with two other existing road extraction methods, namely S.V., which is also region-based and also uses MP, and line-based OTB, which was especially developed for Pléiades images. The proposed method gave higher completeness, correctness and quality than the S.V. method. Visual inspection showed that the proposed method performed better than the OTB method. However, results could not be compared based on completeness, correctness and quality measures, because OTB produces roads as lines and not as regions. As can be seen from Appendix, the proposed method also performs consistently well on a larger subset of the image. Although we compared our method to the best and closest alternatives for road extraction, further comparison with other existing road-extraction methods could be interesting and induce further improvements.

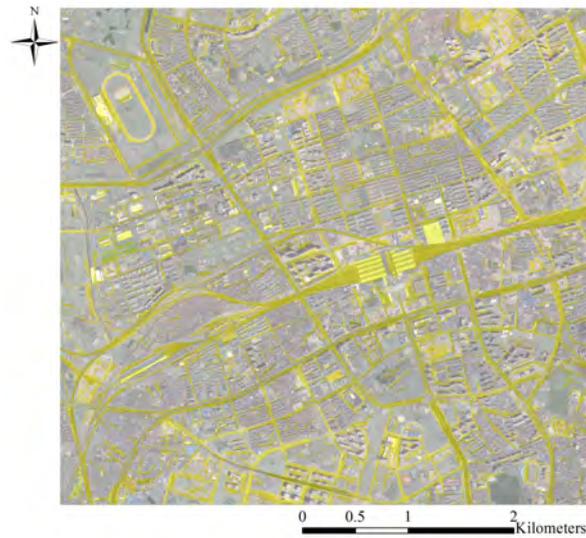
3.6 Conclusions

BPTs have a considerable potential for region-based representation of images. For remote sensing images, so far, the BPTs have been mainly used for image segmentation, based mainly on low-level spectral, textural and geometrical information. We investigated the use of BPTs for region-based urban road extraction from VHR images, integrating geometrical and structural information of roads. Based upon fuzzy logic, different types of information can be effectively fused for constructing a BPT, in which small regions are grouped to form larger and more meaningful road-like regions. It has been known that vegetation, shadows and buildings affect the road extraction from VHR images, particularly in urban environments. We investigated to use a set of automatic thresholding methods to preliminarily remove vegetation, buildings and building shadows to obtain a RoadROI, upon which the proposed road extraction was conducted. The experimental results showed that the proposed road extraction outperformed two other existing methods over the Wuhan and Enschede study areas using Pléiades and Quickbird VHR images respectively. We conclude that the proposed method can provide an alternative means for road extraction over dense urban areas from VHR images, including varying densities of roads, buildings adjacent to roads and partial covering of roads by vehicles and vegetation.

Appendix



(a) Pléiades image



(b) Extracted roads

Figure 3A.1: (a) large Pléiades image with 20000×20000 pixels. (b) extracted roads (yellow) by the proposed method.

Urban land use extraction from Very High Resolution remote sensing imagery using a Bayesian network

4

This chapter is based on the published paper: Li, M., Stein, A., Bijker, W., Zhan, Q., 2016. Urban land use extraction from Very High Resolution remote sensing imagery using a Bayesian network. *ISPRS Journal of Photogrammetry and Remote Sensing* 122, 192–205.

Abstract

Urban land use extraction from Very High Resolution (VHR) remote sensing images is important in many applications. This study explores a novel way to characterize the spatial arrangement of land cover features, and to integrate it with commonly used land use indicators. Characterization is done based upon building objects, taking their functional properties into account. We categorize the objects to a set of building types according to their geometrical, morphological, and contextual attributes. The spatial arrangement is characterized by quantifying the distribution of building types within a land use unit. Moreover, a set of existing land use indicators primarily based upon the coverage ratio and density of land cover features is investigated. A Bayesian network integrates the spatial arrangement and land use indicators, by which the urban land use is inferred. We applied urban land use extraction to a Pléiades VHR image over the city of Wuhan, China. Our results showed that integrating the spatial arrangement significantly improved the accuracy of urban land use extraction as compared with using land use indicators alone. Moreover, the Bayesian network method produced results comparable to other commonly used classifiers. We conclude that the proposed characterization of spatial arrangement and Bayesian network integration is effective for urban land use extraction from VHR images.

Keywords: Urban land use, very high resolution, spatial arrangement characterization, building types, Bayesian network.

4.1 Introduction

Urban land use information plays an important role in many urban-related applications (Patino and Duque, 2013; Hu and Wang, 2013). Remote sensing images have the potential of extracting urban land use and monitoring its changes at local, regional, and national levels (Banzhaf and Netzband, 2011). Particularly, at the local urban level, the growing availability of very high resolution (VHR) remote sensing images, e.g., from QuickBird, GeoEye, WorldView and Pléiades satellites, has resulted in an increase in extracting urban land use with fine spatial detail (Pacifici et al., 2009; Comber et al., 2012; Voltersen et al., 2014). Remote sensing images record the physical properties of the earth surface, i.e. land cover, whereas land use refers to the corresponding functional aspects, i.e. how land cover is used by humans. Extracting urban land use from VHR images is more challenging than land cover extraction, because human activities cannot be directly inferred by interpreting the tone, texture or shapes of image features. For example, a homogenous land use unit (like a street block) may be comprised of a set of different land cover features like vegetation, buildings, and other built-up objects. The composition of land cover features has been widely used for land use classification from low or medium resolution remote sensing images at regional or national levels. It however fails to effectively characterize land use from VHR images at the local level. For example, both high-density residential and commercial land use units have a large coverage of buildings and a low coverage of vegetation. The types and arrangement of land cover features may also vary between land use classes. Therefore, it is crucial to explore an effective way to characterize the arrangement of land cover features for urban land use extraction from VHR images.

Traditional methods for urban land use extraction from remote sensing images rely on quantifying land use indicators (also called spatial metrics) at well-defined land use units, such as grids and street blocks (Herold et al., 2003; Hu and Wang, 2013; Lowry and Lowry, 2014; Novack et al., 2014). Land use indicators that mainly consider the coverage ratio and density of land cover features are commonly used but fail to effectively characterize land use in complex urban areas, and often lead to poor extraction results. For example, a residential area may have buildings with similar size and shape, and a regular spatial layout, but differs from a non-residential area in terms of the spatial layout of land cover features. Therefore, the spatial layout, i.e. spatial arrangement, can be used as a good proxy to differentiate land use classes. Recent studies on urban land use extraction have also shown that the use of spatial arrangement of land cover features can improve the extraction performance (Zhan et al., 2002; van der Kwast et al., 2011; Comber et al., 2012; Vaduva et al., 2013; Walde et al., 2014; Zhang and Du, 2015). Among those, graph-based methods have been regularly used for object-based land use extraction. Those construct a planar graph that characterize the spatial arrangement of land cover features (Barnsley and Barr, 1997; Comber et al., 2012). By doing so, it is challenging to construct, partition, and measure a graph (or sub graphs). In the literature, the extraction of urban structure types has been investigated using VHR images and other data sources such

as a digital surface model (DSM), a digital landscape model (DLM), and LiDAR data (Banzhaf and Hofer, 2008; Heiden et al., 2012; Walde et al., 2014; Voltersen et al., 2014). The urban structure types are defined according to the structural properties including spatial arrangement of image regions. They generally do not refer to the functional properties, i.e. land use, of image regions. Urban structure types may be used as spatial indicators for urban land use extraction. In addition, urban scene classification has drawn increasing attention in remote sensing studies, in particular to VHR remote sensing (Yang and Newsam, 2010; Vaduva et al., 2013; Castelluccio et al., 2015; Zhao et al., 2016). Similar to natural scene classification in computer vision, an urban scene refers to a subset of an image, and is defined according to its semantic interpretation, e.g. an airport, a parking lot, or a sparse residential. The spatial arrangement information based on local features is important for classifying urban scenes (Cheriyadat, 2014; Chen and Tian, 2014). In scene classification, the local features are usually generated from low-level image descriptors, e.g. scale-invariant feature transform (SIFT) in Cheriyadat (2014) and local binary patterns (LBP) in Li et al. (2015b). They however do not provide an explicit interpretation for a specific class of urban scenes (or urban land use).

This paper explores a novel way to characterize spatial arrangement and extract urban land use from VHR images. At the local urban level, a VHR image is dominated by buildings. We therefore characterize the spatial arrangement of land cover features based upon building objects obtained from VHR images. To do so, we define and categorize building objects into a set of building types according to their geometrical, morphological and contextual properties. The distribution of building types is then quantified to represent the spatial arrangement information of a land use unit. We use a Bayesian network to model spatial arrangement information and land use indicators for urban land use extraction.

The remainder of this paper is organized as follows: Section 4.2 describes the study area and data, and Section 4.3 illustrates the proposed method for urban land use extraction. Section 4.4 presents experimental results and related analysis, followed by discussion in Section 4.5 and the conclusions of this research in Section 4.6.

4.2 Study area and data

The study area is an urban area located in Wuhan, the capital of Hubei province, being one of the most populous cities in China. It has a population of more than 10 million inhabitants, which still increases, resulting in rapid land use changes, particularly in newly built urban zones. For this study area, we acquired a subset of a Pléiades-1B image recorded on 11 July 2013, with four multispectral bands of 2 m spatial resolution and a panchromatic band of 0.5 m, covering an area of 25 km². This 0.5 m panchromatic band was resampled from the 0.7 m raw panchromatic band by the image provider using spline resampling. The multispectral bands were fused with the panchromatic band using the Gram-Schmidt pan-sharpening method provided

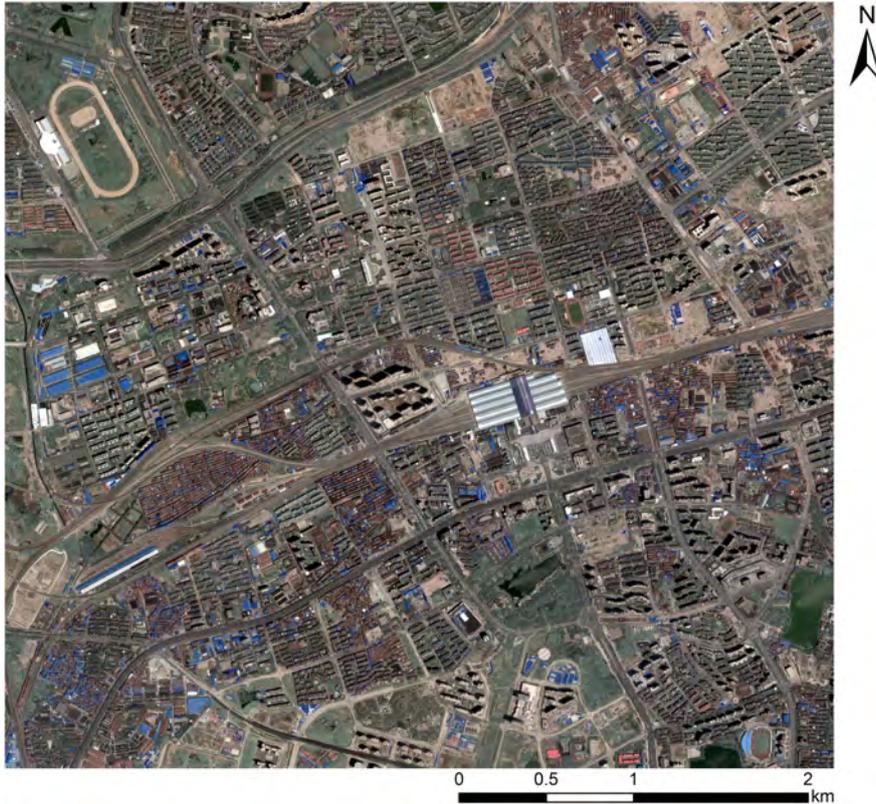


Figure 4.1: True color composite of the pansharpened Pléiades image of the study area in Wuhan, China.

by ENVI version 5.0 (Exelis Visual Information Solutions, Boulder, Colorado). Figure 4.1 gives an overview of the pansharpened Pléiades image in a true color composite. In addition to the VHR image, we acquired an existing urban land use map (around 2013) in shapefile format fully covering this study area from the local department of urban planning. Underlying this map is a detailed classification system of 45 land use classes, among which 38 classes can be found in our study area (Table 4A.1 in Appendix). This land use map was used to 1) define the urban land use classification system for the study area in this paper, 2) partition the study image into homogenous units for land use extraction, and 3) to evaluate the accuracy of urban land use extraction from the study image.

4.3 Methods

The workflow of the proposed urban land use extraction from VHR imagery consists of (1) urban land cover classification, (2) urban building types and land use definition, (3) urban building type classification, (4) spatial ar-

4. Urban land use extraction using a Bayesian network from VHR images

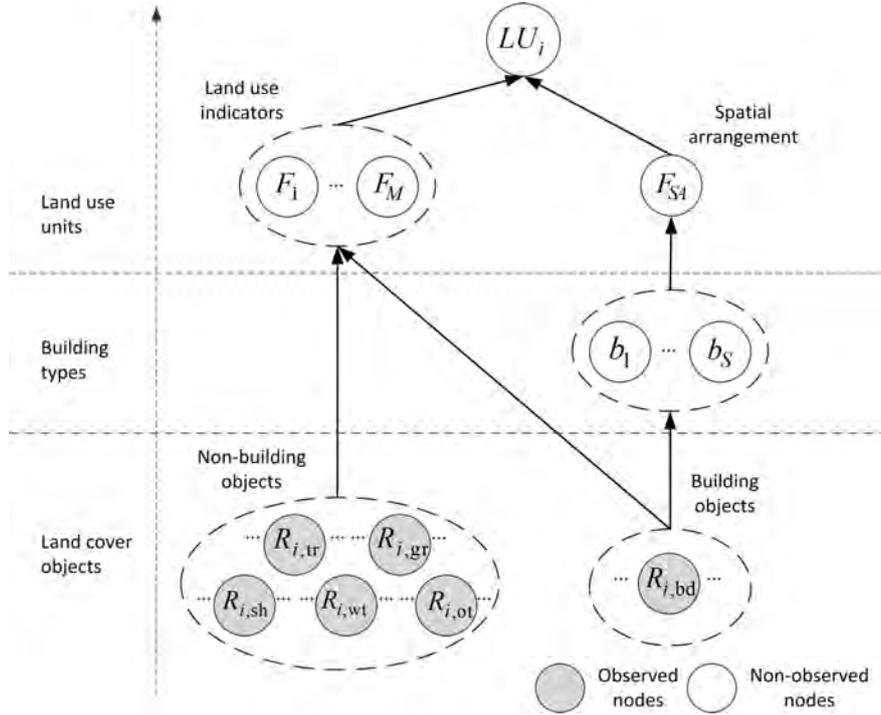


Figure 4.2: The proposed Bayesian network model for extracting land uses from land covers obtained from VHR images. The root node LU_i refers to a land use unit. The group of nodes $\{F_1, \dots, F_M\}$ refer to the used land use indicators. The node F_{SA} refers to the spatial arrangement of a land use unit, and consists of a group of nodes $\{b_1, \dots, b_S\}$ associated with classified building types within this unit. The nodes $R_{i,bd}$, $R_{i,tr}$, $R_{i,gr}$, $R_{i,sh}$, $R_{i,wt}$, $R_{i,ot}$ denote the land cover objects of building (including *dark*, *gray*, *brick-color*, *blue*, and *bright roofs*), *tree*, *grass*, *shadow*, *water*, and *others*, respectively. The non-building objects are used to calculate the value of land use indicators.

rangement characterization, and (5) urban land use extraction by Bayesian network, see Figure 4.2. The evaluation of urban land use extraction is then given. In the following sections, we will describe this in detail.

4.3.1 Urban land cover classification

For classifying urban land cover, we used an object-based image analysis method, for which we defined 11 classes: *grass*, *tree*, *shadow*, *water*, *bare soil*, *dark roof*, *gray roof*, *brick-color roof*, *blue roof*, *bright roof* and *others*. All the *dark*, *gray*, *brick-color*, *blue*, and *bright roofs* were generally considered as buildings. To segment a VHR image into objects, the multiresolution segmentation method in eCognition software was used (Benz et al., 2004). We set the parameters based upon expert experience, i.e. a trial-and-error

Table 4.1: Features used for classifying urban land cover (Li et al., 2015a).

| | | | |
|-----------------------|------------------|-----------------|---------------|
| Spectral | Mean | StdDev | Brightness |
| | Max_Diff | NDVI | NDWI |
| Textual (GLCM) | Homogeneity | Contrast | Dissimilarity |
| | Entropy | Ang. 2nd moment | Mean |
| | StdDev | Correlation | |
| Geometrical | Area | Perimeter | Length/With |
| | Border Index | Compactness | Shape Index |
| | Roundness | Rectangular fit | Elliptic fit |
| Spatial | Directional mean | | |

method, because the study image covers a large and complex urban area, where many buildings are found with diverse materials, sizes, and shapes.

Furthermore, we produced an image segmentation with a slight over-segmentation rather than under-segmentation in order to maintain the accuracy of land cover classification (Myint et al., 2011). For each image object, which is a segment from image segmentation, we calculated a set of spectral, geometrical, textual, and spatial features for classification (Table 4.1). Particularly, we calculated a directional feature based upon building shadow to improve the separability of building objects from other man-made objects (Li et al., 2015a, 2016a). A directional feature quantifies the directional relationship between building and shadow, where the directional relationship is modeled by a fuzzy landscape of building shadow. A support vector machine (SVM) classifier was then used for classification (Chang and Lin, 2011). For the SVM classifier, the radial basis function (RBF) was used for SVM kernel function, and the grid search algorithm was used for determining SVM parameters, i.e. penalty and kernel parameters (Hsu et al., 2010).

4.3.2 Urban building types and land use definition

Buildings can be used for many different purposes, for instance for residential, commercial, or public services use. Based upon VHR images, it is hard to accurately interpret a building object into its real use. But we can categorize building objects into a set of types according to their shapes, sizes, volumes, and contextual relationships with surrounding objects. For example, Wurm et al. (2016) investigated building type classification for several German cities, based upon building models obtained from real estate cadastral building footprints and a detailed DSM at a level of detail 1 (LoD1) (OGC, 2012). A detailed DSM, however, is not always available. Moreover, building types may vary between different cities or countries. For example, some of the building types investigated by Wurm et al. (2016) are rarely found in a Chinese city. By observing the characteristics in our study area, we defined five building types according to their geometrical, morphological and contextual properties. These are *high-density buildings*, *apartment-type buildings*, *complex buildings*, *factory-type buildings*, and *detached buildings*. The land use classes are *low-density residential*, *high-density residential*, *commercial*, *public management and services*, *green space and entertainment*, *redeveloping land*, *transportation*, and *industrial and warehouses*. The defined building types and land use classes are depicted in Figure 4.3, and are described

4. Urban land use extraction using a Bayesian network from VHR images



Figure 4.3: Examples of different building types and land use classes in Wuhan, China, depicted by the study image.

as follows:

- *High-density residential* (Figure 4.3a) areas are used for living, mainly for low income and temporary inhabitants. These areas show a high density of buildings, and a small coverage of open space and vegetation. Buildings within these areas are presented to be adjacent and clustered densely, and are therefore hard to be delineated individually. We define these buildings as *high-density buildings* (B1 of Figure 4.3a).
- *Low-density residential* (Figure 4.3b) areas are also used for living, but mainly for high income and permanent residents. They are normally identified by regularly arranged apartments. These areas have a lower density of buildings, and a larger coverage of open space and vegetation, as compared with *High-density residential* areas. Buildings with similar shapes and sizes (i.e. apartments) are commonly found in areas of this land use class. We define these buildings as *apartment-type buildings* (B2 of Figure 4.3b).

- *Commercial* (Figure 4.3c) areas are used for business and commercial services. These areas are covered by a large proportion of built-up, but little vegetation, and are usually identified by *complex buildings* (B3 of Figure 4.3c) that have large sizes, complex shapes, and are adhere to open space especially for parking lots.
- *Public management and services* (Figure 4.3d) areas are used for government agencies, public service organizations, and science and education. These areas are characterized by relatively low density of buildings, and relatively high coverage of open space and vegetation. These land use areas also contain *complex buildings* for public services. Besides, small buildings which used mainly as supplementary facilities are sparsely distributed in these areas. These buildings are referred to as *detached buildings* (B4 of Figure 4.3d).
- *Green space and entertainment* (Figure 4.3e) areas represent park green space, square green space, green areas for environmental protection, and agricultural and forestry land. The areas are characterized by high coverage of natural land covers, but low for built-up.
- *Transportation* (Figure 4.3f) are used for transportation purpose, and specified by roads, railways, and the corresponding stations.
- *Redeveloping land* (Figure 4.3g) refers to construction sites or demolition areas, showing a high coverage of bare soil and a low coverage of built-up. To be strict, these areas represent a temporal stage of land development, and will be redeveloped to same or different functions with respect to their original land use. Here, we define these areas as a separate land use class.
- *Industrial and warehouses* (Figure 4.3h) areas are used for industries and warehouses. These areas have buildings with a relatively high coverage rates, however have other land cover features with various coverage rates. Within such a land use area, buildings that have regular shapes, various sizes, and low height, are commonly found to being factories. Here we define these buildings as *factory-type buildings* (B5 of Figure 4.3d). Particularly in China, *factory-type buildings* are commonly painted in blue color, corresponding to steel structured factory buildings in architecture.

4.3.3 Building type classification

We exploit the geometrical, morphological, and contextual features of building objects for building type classification (Table. 4.2). The building objects refer to connected components of a classified building image, where pixels equal to 1 refer to building pixels, and equal to 0 otherwise. Particularly, to characterize the building type of *high-density buildings*, we derive a morphological feature to measure their structural properties. Here, we calculate the mean of morphological profiles (MP) created by applying a series of path openings on the pixels of a building object (Li et al., 2016a). It quantifies the structural heterogeneity of this image object. Here, a path opening op-

4. Urban land use extraction using a Bayesian network from VHR images

Table 4.2: Features used for classifying building types.

| Features | | | Literature |
|---|-----------------|-------------|-------------------|
| Geometrical features | | | |
| Area | Perimeter | Length/With | Li et al. (2015a) |
| Border Index | Compactness | Shape Index | |
| Roundness | Rectangular fit | Solidity | |
| Morphological features | | | |
| Mean of morphological profiles | | | Li et al. (2016a) |
| Land cover related features | | | |
| Coverage ratio of blue roofs. | | | |
| Contextual features | | | |
| Pairwise similarity, measures to what degree an image object is similar to its neighboring objects. | | | |

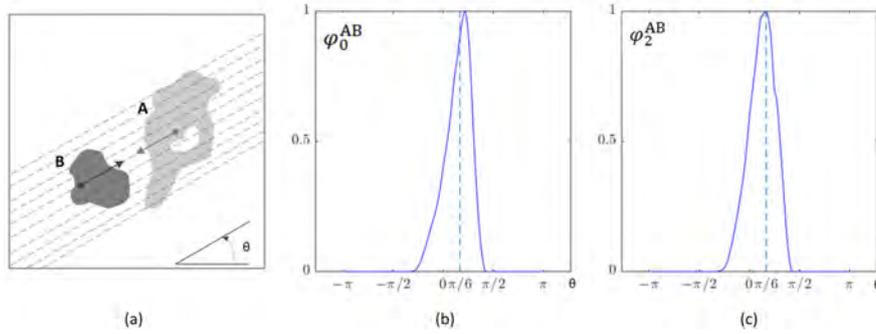


Figure 4.4: Example of an F-Histogram. (a) A pair of image objects A (argument) and B (referent). (b) and (c) are the constant and gravitational F-Histograms of pairwise objects (A, B), respectively. For any direction θ of the referent object B, the value of φ_0^{AB} (resp. φ_2^{AB}) is seen as the sum of elementary forces with respect to the constant (resp. gravitational) representation along direction θ . This value is thus used to quantify the degree of truth of the proposition that object A is in direction θ of B. The F-Histograms in (b) and (c) are normalized by the corresponding largest values.

eration does not need to specify the shape and size of a structural element which is needed for a traditional opening operation.

To characterize the building type of *apartment-type buildings*, we construct a spatial signature, namely pairwise similarity measure, based upon a histogram of forces (F-Histogram). It measures the degree to which an image object is geometrically similar to its neighboring objects. The use of F-Histogram is motivated by previous research (Shyu et al., 2007; Buck et al., 2013; Vaduva et al., 2013), upon which we improve the description of the spatial signature for comparing the shape similarity between a pair of image objects. Like the histogram of angles (Bloch, 2005), an F-Histogram is initially developed to quantify the directional relationships between a pair of image objects (Matsakis and Wendling, 1999; Matsakis et al., 2004). A directional relationship between image objects, and the theoretical founda-

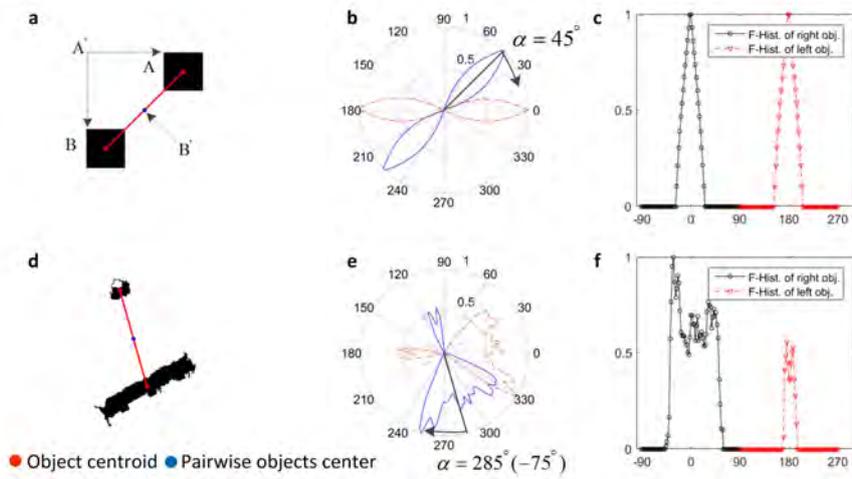


Figure 4.5: Example of the proposed spatial signature for different pairs of image objects. (a) denotes a pair of regular image objects with similar shape. (b) displays the original constant F-Histogram, normalized to $[0,1]$, and the proposed rotationally invariant spatial signatures for (a). (c) decomposes the proposed F-Histogram into two sub histograms corresponding to the two image objects. (d-f) refer to the case for a pair of irregular image objects with different shape.

tions of a F-Histogram are described in detail by Matsakis and Wendling (1999); Matsakis et al. (2004).

Given a pair of image objects A (argument) and B (referent), for every direction $\theta \in [-\pi, \pi]$, we calculate the sum of elementary forces φ_r^{AB} acting between A and B (Figure 4.4). The value of φ_r^{AB} reflects the degree of the support for the proposition that A is in the direction θ of B, where the r represents the magnitude of the elementary forces between points. If r equals 0, we obtain a constant F-Histogram (φ_0^{AB}), which provides a global perspective, independent of the distance between A and B. If r equals 2, we obtain a gravitational F-Histogram (φ_2^{AB}), which provides a local perspective, more sensitive to nearby points, but independent of global scale.

Based upon the F-Histogram, we define a spatial signature for measuring the geometrical properties, in particular the shape similarity, between a pair of image objects. We consider both objects A and B as the new argument A' (Figure 4.5). Then we create a synthetic point as the new referent B' , which is located in the center of objects A and B (see the blue dot in Figure 4.5a). Let (x_A, y_A) and (x_B, y_B) be the x, y locations of the centroids of objects A and B respectively. The location $(x_{B'}, y_{B'})$ of the B' is then equal to $(\frac{x_A+x_B}{2}, \frac{y_A+y_B}{2})$ (Figure 4.5). Here, we obtain the principal axis, for example with an angle α , of the pairwise objects with an angle, and then rotate the original F-Histogram (i.e. blue F-Histograms in Figures 4.5b and e) for α degree in a clockwise direction to achieve a rotationally invariant spatial signature (i.e. red F-Histograms in Figure 4.5)

(Shyu et al., 2007; Vaduva et al., 2013). In addition, we calculate a constant F-Histogram for the proposed spatial signature. Given a spatial signature of a pair of image objects, the obtained F-Histogram can be decomposed into two sub histograms with respect to the two objects (Figures. 4.5c and f). These two sub histograms can be used to compare the shape similarity between the two image objects, by means of histogram comparison. Matsakis et al. (2004) gave three measures for histogram comparison, namely Tversky, Pappis and Cross-correlation measures. In this paper, we use the Cross-correlation measure to compare the shape similarity between a pair of objects using spatial signature (Matsakis et al., 2004). Let h_1 and h_2 be the two sub histograms with respect to the two image objects A and B, the pairwise similarity $\mu_C(A, B)$ between objects A and B is thus defined as

$$\mu_C(A, B) = 1 - \frac{\sum_{\theta} h_1(\theta)h_2(\theta)}{\sqrt{\sum_{\theta} h_1^2(\theta)}\sqrt{\sum_{\theta} h_2^2(\theta)}}, \quad (4.1)$$

where μ_C has a value range $(0, 1]$. The apartment-type buildings are therefore expected to exhibit a high value of μ_C , which can be used as a criterion for extraction.

Last, we use a SVM classifier to classify building objects into five types based upon the above-mentioned geometrical, morphological and contextual features. It is noted that we produce a probabilistic output from SVM classification. The readers are referred to Lin et al. (2007) for details.

4.3.4 Spatial arrangement characterization

The spatial arrangement is characterized as a distribution over a set of classified building types. Given an image with N land use units $LU_i, i = 1, \dots, N$. For each land use unit LU_i , let $b_j, j = 1, \dots, S$ be a variable associated with building objects of a specific type, and S be the number of building types. The variable b_j , in this paper, is calculated as the proportion of building pixels belonging to j th building type within this land use unit LU_i . Accordingly, the spatial arrangement F_{SA} of LU_i is defined as a feature vector, $F_{SA} = [b_1, \dots, b_S]^T$. Since the building type classification may refer to attribute uncertainties, the spatial arrangement F_{SA} can be defined to include the associated uncertainties, i.e. $F_{SA} = [w_1 \cdot b_1, \dots, w_S \cdot b_S]^T$. The parameter w_j is defined as,

$$w_j = \sum_{i=1}^{i=n} \frac{1}{n_j} \cdot p(x_i = b_j) \cdot I(x_i), j = 1, \dots, S, \quad (4.2)$$

where x_i refers to an image pixel within a land use unit, n is the number of pixels within the unit, $p(x_i = b_j)$ is the probability that pixel x_i be classified into j th building type having the number of pixels n_j , and $I(x_i)$ is an indicator function, i.e. $I(x_i) = 1$ if x_i is a building pixel, otherwise $I(x_i) = 0$. In our study we put $S = 5$. The variables F_{SA} are used in a Bayesian network to extract land use from land cover.

Table 4.3: Land use indicators used for characterizing land use. These indicators have been investigated in Voltersen et al. (2014). They are calculated based upon land use units.

| Indicators | Description |
|--------------------|---|
| Geometrical | Area |
| | Perimeter |
| | Compactness |
| | Shape index |
| Land cover related | The coverage ratio of a specific land cover class |
| | The density of a specific land cover class |
| Landscape metrics | Fractal dimension |
| | Landscape shape index |
| | Shannon's diversity index |

4.3.5 Urban land use extraction by Bayesian network

A Bayesian network is a probabilistic graphical model that encodes joint probability distribution over a set of random variables (Nielsen and Jensen, 2007). It is a directed acyclic graph (DAG), where each node represents a random variable (or a group of random variables), and the links express probabilistic relationships between these variables (Bishop, 2006). The graph then captures the way in which the joint distribution over all of the random variables can be decomposed into a product of factors each depending only on a subset of the variables. Thus, for a Bayesian network with K nodes, the joint distribution is given by

$$p(\mathbf{x}) = \prod_{k=1}^K p(x_k | pa_k) \quad (4.3)$$

where x_k is a random variable, pa_k denotes the set of parents of x_k , and $\mathbf{x} = \{x_1, \dots, x_K\}$.

4.3.5.1 Construct Bayesian network structure

The structure of the Bayesian network is based upon expert knowledge on urban land use extraction. It integrates commonly used land use indicators and the spatial arrangement of land cover features. Figure 4.2 gives the structure of the proposed Bayesian network model for extracting land use from a VHR image. The root node LU_i refers to a land use unit. The group of nodes $\{F_1, \dots, F_M\}$ refers to the land use indicators (Table. 4.3), and the node F_{SA} refers to the spatial arrangement of a land use unit. It consists of a group of nodes $\{b_1, \dots, b_S\}$ or $\{w_1 \cdot b_1, \dots, w_S \cdot b_S\}$ associated with classified building types within this unit. The nodes $R_{i,bd}$, $R_{i,tr}$, $R_{i,gr}$, $R_{i,sh}$, $R_{i,wt}$, $R_{i,ot}$ denote land cover objects of building (including *dark*, *gray*, *brick-color*, *blue*, and *bright roofs*), *tree*, *grass*, *shadow*, *water*, and *others*, respectively.

To formalize the problem of urban land use extraction, we assumed that we have a set of variables $F = \{F_1, \dots, F_M, F_{SA}\}$, i.e. the features of a land use unit, and a class variable $C = \{C_1, \dots, C_K\}$, where C_K refers to the states of class C , and K is the number of land use classes. More spe-

4. Urban land use extraction using a Bayesian network from VHR images

cifically, the features F_1, \dots, F_M represent M land use indicators described in Table 4.3, and F_{SA} refers to the spatial arrangement of a land use unit. Based upon our BN, for each unlabeled land use unit, we assigned this land use unit to a category with maximum a posteriori (MAP) likelihood, given the input evidence, corresponding to the calculated F , of this land use unit. The land use class of an unlabeled land use unit is thus inferred by

$$C^* = \underset{C_i}{\operatorname{argmax}} p(C_i|F), i = 1, \dots, K. \quad (4.4)$$

By adopting Bayes' theorem

$$p(C_i|F) = \frac{p(C_i)p(F|C_i)}{p(F)}, \quad (4.5)$$

we obtain

$$C^* = \underset{C_i}{\operatorname{argmax}} p(C_i)p(F|C_i), i = 1, \dots, K, \quad (4.6)$$

where $p(F)$ is the same for all the possible states of C_i .

4.3.5.2 Learn parameters for Bayesian network

The $p(C_i)$ models the prior distribution of different land use classes. In this paper, we do not consider specific priors for land use extraction, and use a flat prior. The conventional methods for estimating the conditional probability $p(F|C_i)$ are 1) to first discretize a continuous variable $F_i \in F$ into a categorical variable, and then calculate the frequencies of each state of this categorical variable using a large number of training samples, or 2) to assume that F follows a multivariate Gaussian distribution, for which many training samples are needed to calculate its mean and standard deviation (Nielsen and Jensen, 2007). To avoid such discretization, Gaussian assumption, and the need of large training samples, we apply a kernel method to estimate the conditional probability $p(F|C_i)$ (Lin et al., 2007; Cheng and Wang, 2010). For both features referring to land use indicators and spatial arrangement, we use a set of labeled land use units to train a SVM classifier. According to Lin et al. (2007), we thus obtain a probabilistic output, used as conditional probabilities. Since the spatial arrangement F_{SA} is characterized over the distribution of building types, we use an intersection kernel for histogram-based comparison (Maji et al., 2013). Last, based upon our Bayesian network, we classify the unlabeled land use units according to Equation 4.6.

4.3.6 Evaluation of urban land use extraction

We compare the proposed Bayesian network method that incorporates spatial arrangement and land use indicators, labeled as BN ($F_M + F_{wSA}$) when the associated uncertainties are considered and BN ($F_M + F_{SA}$) otherwise, with that using land use indicators alone, labeled as BN (F_M). In addition, we compare the proposed method with two other existing classifiers,

i.e. SVM and random forest (RF). These two classifiers have shown good performance in many classification problems. Both spatial arrangement that characterized considering uncertainties and land use indicators are used for urban land use extractions by SVM and RF, labeled as SVM ($F_M + F_{wSA}$) and RF ($F_M + F_{wSA}$) respectively. We assess the extracted urban land use by means of confusion matrix. Moreover, we conduct a McNemar test to investigate the significance of difference of every comparison (Foody, 2004).

4.4 Results

4.4.1 Urban land cover classification

For image segmentation, we performed a series of image segmentations using eCognition software by varying the scale parameters from 80 to 150, while other parameters were set as default. By visual inspection, we chose the image segmentation result produced from the scale parameter of 120. In total, the Pléiades VHR image was segmented to 124,389 objects, from which we selected 1213 sample objects for the 11 defined land cover classes. The sample objects were randomly divided into training and testing datasets, corresponding to 550 and 663 sample objects respectively. Specifically, we randomly selected 50 sample objects for training each land cover class, and the remaining sample objects for testing classification performance. A SVM classifier was used to implement the land cover classification. Figure 4.6 shows the results of the classified urban land cover from the study image. To evaluate the accuracy of land cover classification, the testing dataset was used to calculate a confusion matrix, see Table 4.4. This table shows an overall accuracy of 90.10%, and the corresponding kappa coefficient of 0.8898. Except for the land cover class *others*, all the land cover classes were classified with user accuracies of more than 80%, particularly, the land cover of *grass*, *tree*, *shadow*, *dark roof*, *bright roof*, and *blue roof* have user accuracies more than 90%.

4.4.2 Building type classification

To calculate the mean of morphological profiles for characterizing *high-density residential building* type, we created morphological profiles from a series of path openings. Here, the longest length of the used path openings should be large enough to capture the structure of the largest building object. Hence, we gave a value of 400 pixels (equal to 200 m) that was big enough to capture all the possible building objects in this study dataset. To calculate the pairwise similarity feature for characterizing *apartment-type buildings*, we first determined the neighborhood for each building object. Vanegas et al. (2013) investigated several choices to determine a neighborhood for image objects, e.g. a voronoi neighborhood. We used circular neighborhoods of a building object. These were constrained by the boundary of the corresponding land use unit. The value of the radius was empirically determined. A bigger value implies that more objects in a bigger neighborhood are found. The

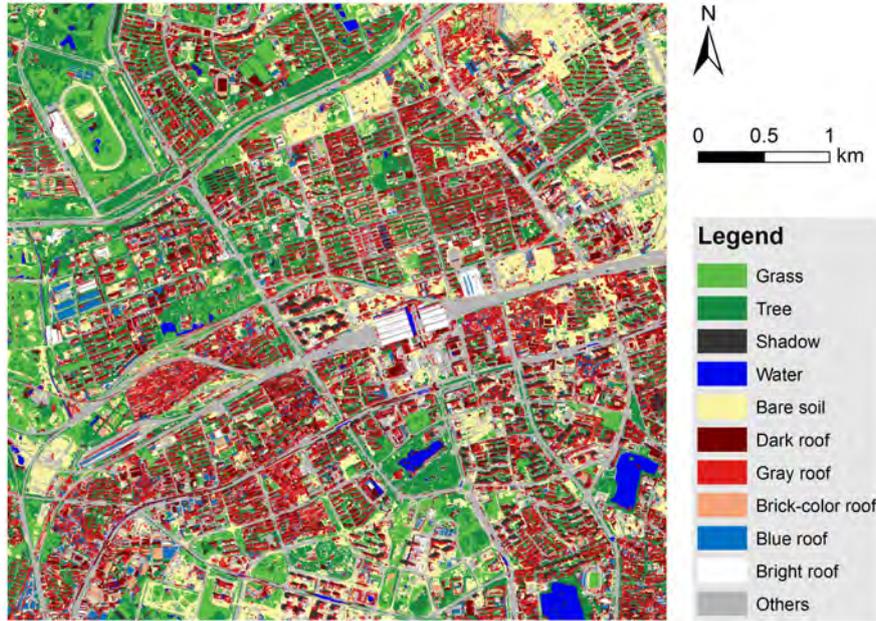


Figure 4.6: Urban land cover classified from the Pléiades VHR image.

radius here was empirically set to 200 pixels, equal to 100 m on ground, to compromise the neighborhood size and computation time.

For building type classification, we also implemented a SVM classification, for which 50 building objects were selected for training each building type. Here we did not consider those building objects that appeared in transportation land use units for building type classification. To test the performance of the selected training dataset, a 5-fold cross validation was conducted. It produced an accuracy of 85.29%. Figure 4.7 shows the results of the classified building types, and visualizes the associated uncertainties based on the probabilistic output of SVM classification. By comparing the Figures. 4.1 and 4.7, we can find that 1) *high-density buildings* are mainly in *high-density residential* areas, 2) *apartment-buildings* are regularly found in *low-density residential* areas, 3) *complex buildings* are found in *commercial, public management and services*, and *industrial and warehouses* land use areas, 4) *factory-type buildings* are mainly found in *industrial and warehouses* land use areas, and 5) *detached buildings* are distributed in almost every land use class.

4.4.3 Urban land use extraction by Bayesian network

We applied a kernel method, implemented by a SVM classifier that produced a probabilistic output, to calculate the corresponding conditional probabilities. To be specific, the conditional probability $p(F_M|C_i)$ of land use indicators was calculated by the SVM classifier with a RBF kernel, while the conditional probability $p(F_{SA}|C_i)$ of spatial arrangement was calculated by

Table 4.4: Confusion matrix of the urban land cover classification. G, T, S, W, BS, D, GR, BR, B, L, O indicate the land cover of *grass*, *tree*, *shadow*, *water*, *bare soil*, *dark roof*, *brick-color roof*, *blue roof*, *bright roof*, and *others*, respectively. PA and UA refer to the producer and user accuracy of the confusion matrix, respectively.

| Cls. | Ref. | | | | | | | | | | | UA |
|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| | G | T | S | W | BS | D | GR | BR | B | L | O | |
| G | 46 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 97.87 |
| T | 0 | 45 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100.00 |
| S | 0 | 0 | 64 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 96.97 |
| W | 0 | 0 | 3 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 82.61 |
| BS | 1 | 0 | 0 | 0 | 41 | 0 | 0 | 0 | 0 | 0 | 5 | 87.23 |
| D | 0 | 0 | 1 | 1 | 0 | 54 | 2 | 0 | 0 | 0 | 0 | 93.10 |
| GR | 1 | 0 | 0 | 0 | 5 | 4 | 83 | 1 | 0 | 2 | 2 | 84.69 |
| BR | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 43 | 0 | 0 | 0 | 100.00 |
| B | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 42 | 0 | 0 | 97.67 |
| L | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 42 | 2 | 87.50 |
| O | 0 | 0 | 0 | 0 | 16 | 1 | 1 | 3 | 0 | 0 | 67 | 76.14 |
| PA | 95.83 | 97.83 | 94.12 | 95.00 | 66.13 | 88.52 | 94.32 | 91.49 | 95.45 | 95.45 | 85.90 | % |
| Overall accuracy = 90.10 %, kappa coefficient = 0.8898. | | | | | | | | | | | | |

the SVM classifier with a histogram intersection kernel. To train the two SVM classifiers, we selected 540 land use samples based on the study image and official land use map, and randomly divided them into two training and testing datasets, with size of 270 and 270 respectively. The training dataset included 40 samples for each land use class except for the *high-density residential*, for which we chose 30 samples, because, in this study area, the *high-density residential* class covered less area than the other land use classes. Figure 4.8 gives the extracted urban land use from the Pléiades image by the proposed method BN ($F_M + F_{wSA}$).

Figure 4.9 exemplifies the main process of the proposed urban land use extraction from VHR images. For example, Figure 4.9a is an example of extracting *high-density residential* land use. The VHR image within a land use unit, i.e. Figure 4.9a1, was first classified into urban land cover (Figure 4.9a2), from which buildings were categorized into five building types (Figure 4.9a3). The distribution of these five building types within this land use unit was quantified in terms of their proportions, i.e. spatial arrangement in Figure 4.9a4. The spatial arrangement information was then used to characterize and extract land use for this land use unit, shown in Figure 4.9a5. Figure 4.9 also shows that the spatial arrangement between different land use classes are evidently different, particularly to those land use classes having a large coverage of buildings, e.g. *high-density residential*, *low-density*

4. Urban land use extraction using a Bayesian network from VHR images

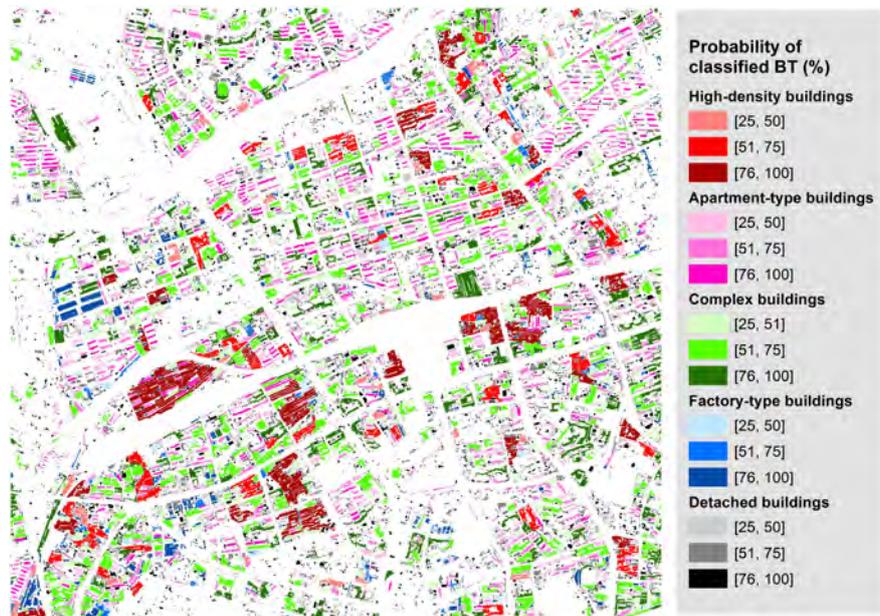


Figure 4.7: Classified building types from the Pléiades VHR image, and visualization of the associated uncertainties.

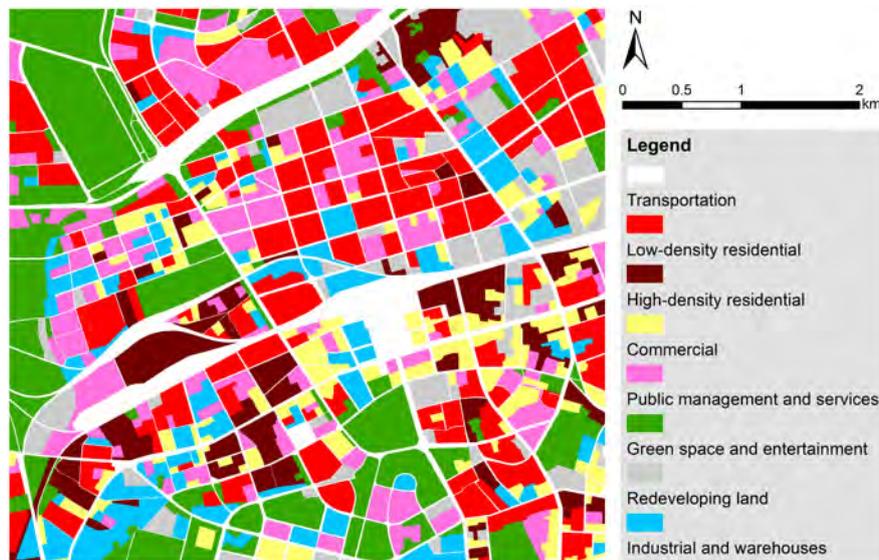


Figure 4.8: Extracted land use from the Pléiades VHR image using the proposed Bayesian network method BN ($F_M + F_{wSA}$).

residential, commercial and industrial and warehouses.

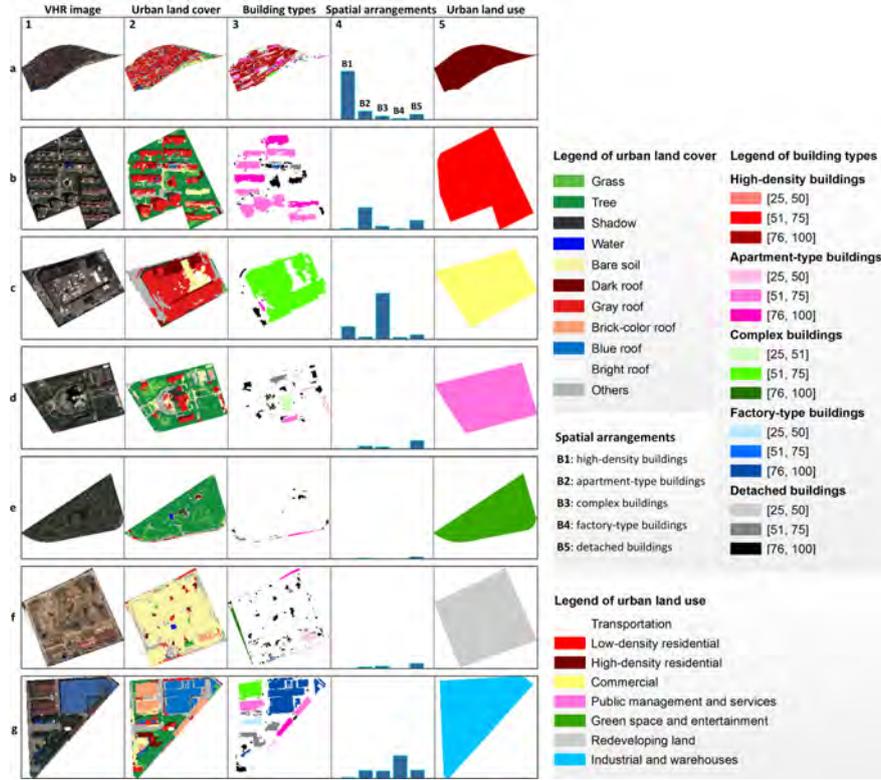


Figure 4.9: Examples demonstrating the main process of the proposed urban land use extraction from a VHR image. Rows (a-g) correspond to the examples of different land use classes. Columns (1-5) show the outputs of each main process starting from a VHR image and ending with an urban land use.

4.4.4 Evaluation of urban land use extraction

A confusion matrix was calculated based on testing dataset to assess the urban land use extraction by BN ($F_M + F_{wSA}$), see Table 4.5. This table shows that the proposed urban land use extraction had an overall accuracy of 83.33% and a kappa coefficient of 0.7959. The extracted *low-density residential* had the highest producer accuracy, equal to 97.30%, while the extracted *high-density residential* had the lowest producer accuracy, equal to 53.85%. Because of the relatively small coverage of *high-density residential* class in this study area, only 13 samples of *high-density residential* were found in the testing dataset. The accuracy of extracted *high-density residential* was thus sensitive to the number of its testing samples.

To investigate the importance of incorporating spatial arrangement information for urban land use extraction, we compared the extraction methods between BN ($F_M + F_{wSA}$) and BN (F_M). The BN (F_M) method extracted the urban land use with an overall accuracy of 69.26% and a kappa

4. Urban land use extraction using a Bayesian network from VHR images

Table 4.5: Confusion matrix of the urban land use extraction by the proposed method BN ($F_M + F_{wSA}$). LR, HR, CM, PM, GS, RE, IW indicate the land use of *low-density residential*, *high-density residential*, *commercial*, *public management and service*, *green space and entertainment*, *redeveloping land*, and *industrial and warehouses*, respectively. PA and UA refer to the producer and user accuracy of the confusion matrix, respectively.

| Cls. | Ref. | | | | | | | |
|------|-------|-------|-------|-------|-------|-------|-------|-------|
| | LR | HR | CM | PM | GS | RE | IM | UA |
| LR | 72 | 3 | 10 | 1 | 0 | 0 | 2 | 81.82 |
| HR | 0 | 7 | 1 | 0 | 0 | 0 | 2 | 70 |
| CM | 0 | 2 | 27 | 1 | 0 | 0 | 1 | 87.10 |
| PM | 2 | 0 | 2 | 30 | 0 | 1 | 3 | 78.95 |
| GS | 0 | 0 | 0 | 1 | 48 | 0 | 0 | 98.00 |
| RE | 0 | 0 | 0 | 1 | 4 | 17 | 1 | 73.91 |
| IW | 0 | 1 | 1 | 5 | 0 | 0 | 24 | 77.42 |
| PA | 97.30 | 53.85 | 65.85 | 76.92 | 92.31 | 94.44 | 72.73 | % |

Overall accuracy = 83.33 %, kappa coefficient = 0.7959.

coefficient of 0.6273. Table 4.6 gives the results of a McNemar test upon the extracted urban land use between BN ($F_M + F_{wSA}$) and BN (F_M). This table shows that these two extractions were significantly different.

We distinguished between spatial arrangement characterizations with and without considering the associated uncertainties of classified building types. Compared with BN ($F_M + F_{wSA}$), the BN ($F_M + F_{SA}$) method produced an urban land use extraction with a slightly lower accuracy, corresponding to an overall accuracy of 81.48% and a kappa coefficient of 0.7731, respectively. Table 4.7 gives the results of a McNemar test upon the extract urban land use between BN ($F_M + F_{wSA}$) and BN ($F_M + F_{SA}$). This table does not indicate a significant different between these two extractions.

We compared the urban land use extraction by BN ($F_M + F_{wSA}$) with two other extractions by SVM ($F_M + F_{wSA}$) and RF ($F_M + F_{wSA}$). The training samples for SVM ($F_M + F_{wSA}$) and RF ($F_M + F_{wSA}$) were the same as those used for calculating the conditional probabilities for a Bayesian network. The SVM ($F_M + F_{wSA}$) classifier had a RBF kernel, and used grid search for determining its parameters. Table 4.8 gives the quantitative evaluation of the urban land use extracted by BN ($F_M + F_{wSA}$), SVM ($F_M + F_{wSA}$), and RF ($F_M + F_{wSA}$) in terms of overall accuracy and kappa coefficient. This table shows that BN ($F_M + F_{wSA}$) produced the highest overall accuracy and kappa coefficient, while SVM ($F_M + F_{wSA}$) and RF ($F_M + F_{wSA}$) methods produced nearly similar accuracies. Furthermore, a McNemar test was conducted to test the statistical significance of the difference between BN ($F_M + F_{wSA}$), SVM ($F_M + F_{wSA}$), and RF ($F_M + F_{wSA}$) extractions, see Table 4.9. This table shows that, although the BN ($F_M + F_{wSA}$) method produced the highest accuracy, there is no significant difference between them.

A sensitivity analysis was conducted to investigate the effect of different land cover classifications on final land use extractions (Figure 4.10). Since we used an object-based method for land cover classification, here we first conducted a series of image segmentations using eCognition, for which the

Table 4.6: Summary of the McNemar test to compare the proportions of correctly extracted land use units between BN ($F_M + F_{wSA}$) and BN (F_M). f_{12} and f_{21} are the number of discordant pairs, i.e. the number of land use units that classifiers classify differently.

| f_{12} | f_{21} | $f_{12} - f_{21}$ | $ z $ | Significant? Level |
|----------|----------|-------------------|-------|-----------------------|
| 54 | 16 | 38 | 4.42 | Yes, 0.1% |

Table 4.7: Summary of the McNemar test to compare the proportions of correctly extracted land use units between BN ($F_M + F_{wSA}$) and BN ($F_M + F_{SA}$). f_{12} and f_{21} are the number of discordant pairs, i.e. the number of land use units that classifiers classify differently.

| f_{12} | f_{21} | $f_{12} - f_{21}$ | $ z $ | Significant? Level |
|----------|----------|-------------------|-------|-----------------------|
| 12 | 7 | 5 | 0.92 | No, 0.1% |

Table 4.8: Comparison of three different urban land use extractions using the proposed BN ($F_M + F_{wSA}$), SVM ($F_M + F_{wSA}$), and RF ($F_M + F_{wSA}$) respectively.

| Classifications | Overall accuracy (%) | Kappa coefficient |
|-------------------------|----------------------|-------------------|
| BN ($F_M + F_{wSA}$) | 83.33 | 0.7959 |
| SVM ($F_M + F_{wSA}$) | 78.89 | 0.7426 |
| RF ($F_M + F_{wSA}$) | 79.63 | 0.7507 |

Table 4.9: Summary of the McNemar test of the extractions using the proposed BN ($F_M + F_{wSA}$), SVM ($F_M + F_{wSA}$), and RF ($F_M + F_{wSA}$) respectively.

| Comparisons | f_{12} | f_{21} | $f_{12} - f_{21}$ | $ z $ | Significant? Level |
|--|----------|----------|-------------------|-------|-----------------------|
| SVM ($F_M + F_{wSA}$) vs. RF ($F_M + F_{wSA}$) | 12 | 14 | -2 | 0.19 | No, 0.1 % |
| SVM ($F_M + F_{wSA}$) vs. BN ($F_M + F_{wSA}$) | 10 | 22 | -12 | 1.94 | No, 0.1 % |
| RF ($F_M + F_{wSA}$) vs. BN ($F_M + F_{wSA}$) | 14 | 24 | -10 | 1.46 | No, 0.1 % |

scale parameter was assigned to values ranging from 60 to 140 with a increment of 20, whereas other parameters were set as default. We labeled an image segmentation for example at scale 60 as L60 for convenience. We then performed a series of urban land cover classifications using SVM, followed by a series of urban land use extractions using the proposed Bayesian network method. As mentioned in section 4.1, we selected 1213 sample image objects for 11 land cover classes at scale 120 (i.e. L120). We mapped these sample image objects at scale L120 into other scales, i.e. L60, L80, L100, and L140, to obtain respective samples at each scale. Particularly, if a sample image object at L120 covers multiple image objects at finer scales, i.e. L60, L80, and L100, only the image object having the largest coverage with this sample image object was selected as the corresponding sample image object at lower scales. By doing so, for each segmentation scale, we selected 1213 sample image objects, which were randomly divided into training (550) and testing (663) datasets for a land cover classification. This random division

4. Urban land use extraction using a Bayesian network from VHR images

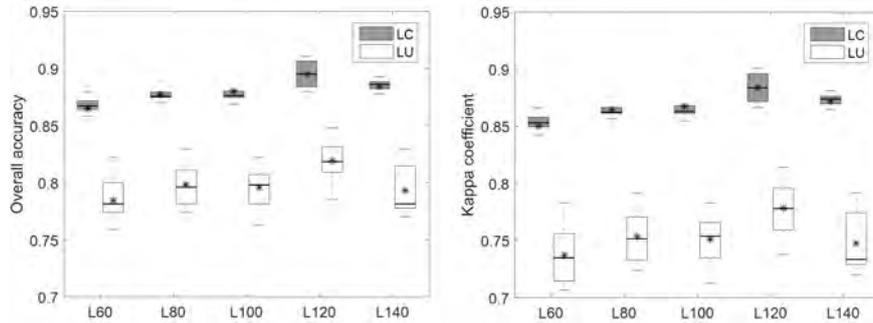


Figure 4.10: Sensitivity analysis of the effect of giving different land cover classifications on final land use extractions by varying image segmentations from scale 60 (i.e. L60) to scale 140 (i.e. L140). The left refers to overall accuracies of urban land cover (LC) and land use (LU) extractions, and the right indicates their kappa coefficients. Each urban land cover or land use extraction run 10 times using different training and testing datasets.

was repeated 10 times so that 10 different land cover classifications were generated. As mentioned in section 4.3, we selected 540 land use samples and randomly divided them into training (270) and testing (270) datasets. This random division was also repeated 10 times to obtain 10 different land use extractions at each scale. The distribution of the overall accuracies and kappa coefficients of these land cover and land use extractions was shown in Figure 4.10. This figure shows that there is a positive correlation between a land cover classification and its subsequent land use extraction. In addition, image segmentation scale L120 produced the highest overall accuracy and kappa coefficient for both land cover and land use extractions, whereas L60 yielded the lowest accuracy. Both land cover and land use extraction have a similar performance at scales L80, L100, and L140 in terms of overall accuracy and kappa coefficient.

4.5 Discussion

This paper provided a method for urban land use extraction from VHR images, in particular exploring a novel way for spatial arrangement characterization. Our findings suggest that the spatial arrangement of land cover features characterized by quantifying the distribution of building types, can effectively improve urban land use extraction. Information on spatial arrangement has been successfully used for scene classification (Cheriyadat, 2014; Chen and Tian, 2014; Zhao et al., 2016). These studies considered the distribution of low-level image features or new features abstracted from low-level image features. We investigated more meaningful geospatial objects, namely building types, for spatial arrangement characterization, rather than using low-level image features such as SIFT features. Thus, the characterized spatial arrangement is able to provide an explicit interpretation for

urban land use classes. Besides spatial arrangement characterization, this paper also contributed a hierarchical representation of urban land use by a tree-shape Bayesian network.

The proposed urban land use extraction relied on classified urban land cover. The sensitivity analysis also shows that there is positive correlation between a land cover classification and its subsequent land use extraction. Many studies on urban land cover classification can be found in the literature (Myint et al., 2011). Since this paper focused on urban land use rather than land cover, we applied a regular object-based method for classifying urban land cover. Other advanced land cover classification methods may improve urban land use extraction. For example, buildings are commonly found with complex spectral and spatial characteristics in densely populated urban areas. Shadow and dark roofs may have similar spectral, textual, and geometrical characteristics, resulting in a high mis-classification rate between them using a regular classification method.

Five building types are defined in this paper. To the best of our knowledge, these five building types have not been investigated in the literature. A recent study on building type classification was given by Wurm et al. (2016). They defined five building types for two cities in Germany, and classified building types based on building models at LoD1 using both 2D and 3D features. Some of those building types, however, can not be found in our study area, e.g. perimeter block development. Differing from that, we defined five building types that are more specific for our study area, and classified based only on 2D features. We further defined eight urban land use classes. Those were modified from an official land use map provided by the local department of urban planning to fit the characteristics of the study area and dataset (Tables 4A.1 and 4A.2 in Appendix). A well-tuned definition of building types and land use classes, for any specific area in the world should depend upon study images and areas. The proposed framework, which starts from urban land cover classification, goes to spatial arrangement characterization, and ends with urban land use extraction, however, can be adopted for any urban area with a well-tuned definition of building types and land use classes.

A pairwise similarity feature is defined based upon the histogram of forces to measure the shape similarity between two image objects. This F-Histogram has been investigated to measure the spatial relationships between two image objects in previous studies (Shyu et al., 2007; Vaduva et al., 2013). We found that a F-Histogram for a pair of image objects can be decomposed into two sub histograms corresponding to each image object. The shape similarity of this pair of image objects can therefore be measured by comparing these two sub histograms. The question is then to determine a proper neighborhood of an image object. We adopted circular neighborhoods of image objects, for which their radius were empirically set to 200 pixels, equal to 100 m. For building objects thus related the spatial relations identify building types, i.e. the importance of the relations decreases with distance. Therefore it is of little benefit to use a large search radius. In addition, Wurm et al. (2016) recommended that the proximity of building objects to surroundings might improve the classification of building types. They however did not use

4. Urban land use extraction using a Bayesian network from VHR images

that information in their work. Here we show that it can be advantageous to do so and hence expand their research.

Effective urban land use extraction also depended upon the choice of suitable land use units. Unsuitable land use units may under- or over- generalize the obtained land use map, particularly with respect to the boundaries between land use classes. We used a well-established urban land use map to segment a VHR image for obtaining homogenous land use units. This differs from other studies on land use, e.g. van der Kwast et al. (2011) using regular grids, Comber et al. (2012) using sub-graphs partitioned from a planar graph, and Hu and Wang (2013) using parcels. An existing urban land use map however may not be available for some areas. A road map, e.g. derived from Open Street Map or from a VHR image (Li et al., 2016a), may then be helpful to obtain basic land use units. In such a way, mixed land use, i.e. a land use unit containing multiple land use classes, might be taken into account.

The strength of the Bayesian network of modeling urban land use was that it was able to fuse different types of knowledge and handle associated uncertainties. Its difficulty is, however, to build a structure (i.e. DAG) and estimate its parameters. We built the structure of our Bayesian network based upon expert knowledge. We assumed that the spatial arrangement variable was independent of variables of land use indicators. This assumption produced acceptable results. Future improvement could be done by considering a dependence relationship of these two types of variables. We used a flat prior for $p(C_i)$, and applied a kernel method to estimate conditional probabilities $p(F|C_i)$. More specifically, we used a RBF kernel and an intersection kernel to estimate the conditional probabilities $p(F_M|C_i)$ and $p(F_{SA}|C_i)$ respectively, where F_M referred to variables of land use indicators, and F_{SA} referred to variable of spatial arrangement. By doing so, no Gaussian assumption or discretization was needed to deal with continuous variables. Our results also confirmed previous findings using kernel methods to estimate conditional probabilities by Lin et al. (2007); Cheng and Wang (2010). Furthermore, the obtained land use extraction results can be served as expert knowledge for extracting land use in other urban areas, or in the same area at different dates for changes detection. This will also lead to our future work.

Finally, we observed that the *high-density residential* land use was extracted with the lowest accuracy. Because of the small coverage of this class in our study area, we used less samples to both estimate parameters (by kernel methods) and validate extracted results than other land use classes. The poor extraction of *high-density residential* might be caused by the imbalance of land use samples in both training and testing datasets. For accuracy assessment, future improvements could be done by implementing and evaluating urban land use extraction for multiple times by varying training and testing datasets.

4.6 Conclusions

This study investigated urban land use extraction from VHR remote sensing images. To infer land use from land cover, we focused on the characterization of spatial arrangement of land cover features, besides of more common land use indicators. We investigated the spatial arrangement characterization via a set of building types defined in this paper. The building types were classified from building objects according to their geometrical, morphological, and contextual properties. A Bayesian network incorporating spatial arrangement information and land use indicators was constructed to represent and extract urban land use. Experimental results showed that spatial arrangement characterized by the proposed method significantly improved the accuracy of extracted urban land use. The proposed Bayesian network method produced results comparable to other commonly used classifiers. Moreover, for Bayesian network modeling, the adopted kernel method was effective to estimate conditional probabilities for variables, when training dataset was limited. We concluded that the proposed spatial arrangement characterization was effective for identifying urban land use, and the proposed Bayesian network method provided an alternative means for urban land use extraction from VHR images.

Appendix

Table 4A.1: Original land use classification system for existing land use map.

| Level I: abbr. | Level I: name | Level II: abbr. | Level II: name |
|-------------------|--------------------------------|--------------------|--|
| R | Residential | R1 | 1st-class residential |
| | | R2 | 2nd-class residential |
| | | R3 | 3rd-class residential |
| I | Industrial | I1 | 1st-class industrial |
| | | I2 | 2nd-class industrial |
| | | I3 | 3rd-class industrial |
| L | Logistics and warehousing | L1 | 1st-class logistics and warehousing |
| | | L2 | *2nd-class logistics and warehousing |
| | | L3 | 3rd-class logistics and warehousing |
| P | Public management and services | P1 | Administrative office |
| | | P2 | Cultural facilities |
| | | P3 | Science and education |
| | | P4 | Sports |
| | | P5 | Medical and health services |
| | | P6 | Social welfare facilities |
| | | P7 | Historical sites |
| C | Commercial | C1 | Commercial |
| | | C2 | Business and finance |
| | | C3 | Recreation |
| | | C4 | Business-net spots for public facilities |
| | | C5 | Other facilities |
| N | Non-construction | N1 | Water |
| | | N2 | Agriculture and forestry |
| | | N3 | Other non-construction land |
| D | Demolition and vacant | D1 | Demolition land |
| | | D2 | *Vacant land |
| G | Green space | G1 | Park green space |
| | | G2 | Environmental protection green area |
| | | G3 | Square green space |
| | | G4 | Productive plantation area |
| H | Others | O1 | *Construction land for rural residents |
| | | O2 | Regional transport facilities |
| | | O3 | Regional public facilities |
| | | O4 | Special use land |
| | | O5 | *Mining sites |
| | | O6 | *Other construction land |
| T | Transportation | T1 | Urban road |
| | | T2 | Urban rail transit |
| | | T3 | Integrated transport hub |
| | | T4 | Transport stations |
| | | T5 | Other transport stations |
| F | Public facilities | F1 | Supply facilities |
| | | F2 | Environment facilities |
| | | F3 | Safety facilities |
| | | F4 | Other public facilities |

* indicates that a class from the existing land use map does not appear in the study area.

Table 4A.2: Land use classification system for this paper.

| Name | Sub-class |
|--------------------------------|--------------|
| High-density residential | R3 |
| Low-density residential | R1, R2 |
| Commercial | C |
| Public management and services | P, F, O3, O4 |
| Green space and entertainment | G, N |
| Transportation | T, O2 |
| Redeveloping land | D1 |
| Industrial and warehouses | I, L1, L3 |

Incorporating open source data for Bayesian classification of urban land use from VHR stereo images

5

This chapter has been submitted to: Li, M., de Beurs, K.M., Stein, A., Bijker, W., 2017. Incorporating open source data for Bayesian classification of urban land use from VHR stereo images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Abstract

This study investigates the incorporation of open source data into a Bayesian classification of urban land use from Very High Resolution (VHR) stereo satellite images. The adopted classification framework starts from urban land cover classification, proceeds to building type characterization, and results in urban land use. For urban land cover classification, a preliminary classification distinguishes trees, grass, and shadow objects using a random forest at a fine segmentation level. Fuzzy decision trees derived from hierarchical Bayesian models separate buildings from other man-made objects at a coarse segmentation level, where an open street map provides prior building information. A Bayesian network classifier combining commonly used land use indicators and spatial arrangement is used for the urban land use classification. The experiments were conducted on GeoEye stereo images over Oklahoma, the United States. Experimental results showed that the urban land use classification using VHR stereo images performed better than that using a monoscopic VHR image, and the integration of open source data improved the final urban land use classification. Our results also confirmed that the adopted urban land use classification framework, developed for a specific urban area, is transferable to other urban areas. The study concludes that incorporating open source data by Bayesian analysis improves urban land use classification. Moreover, a pretrained convolutional neural network fine tuned on the UC Merced land use dataset offers a useful tool to extract additional information for urban land use classification.

Keywords: Urban land use, VHR stereo images, fuzzy decision trees, Bayesian methods, open source data, convolutional neural networks (CNN)

5.1 Introduction

According to UN-Habitat (2016), 54 percent of the world’s population resides in urban areas, and this figure is expected to rise to 66 percent by 2050. Along with rapid population increase we may expect rapid urban expansion, resulting in a surge of new built-up areas for residential, commercial and other urban land use types. Detailed and accurate urban land use information is thus important to study the interaction between humans and the environment and to support sustainable urban development. Increasing attention has been paid to the extraction of urban land use from remote sensing images (Barnsley and Barr, 1997; Herold et al., 2003; van der Kwast et al., 2011; Comber et al., 2012; Voltersen et al., 2014; Li et al., 2016b). Particularly, satellites like GeoEye, Worldview, and Pléiades are providing large volumes of Very High Resolution (VHR) images, in single, stereo or tri-stereo modes, allowing for the extraction of land use information with fine spatial detail (Pesaresi et al., 2013).

Many studies on land use classification using remote sensing images, acquired from various platforms and sensors, can be found (Ünsalan and Boyer, 2011). This paper aims to classify land use over urban areas using VHR images at local level. Typically, land use is distinguished from land cover, where land use refers to the use of land by humans, and is usually composed of different land cover types. Spectral-based features are therefore insufficient to characterize land use classes, and this problem is especially important in urban settings (Zhang et al., 2015). Many studies have indicated that the use of contextual or spatial information is necessary to improve the separability between different land use types (van der Kwast et al., 2011; Li et al., 2016b). Here, we categorize existing relevant studies into grid-based, graph-based, scene-based, and vector-based (including parcel-based) methods based on how available spatial information is used in the land use classification.

Grid-based methods use a regular grid (i.e. a moving window) to define the neighborhood of interest (Eyton, 1993). Contextual information is then derived within this regular neighborhood. These methods were commonly used for medium- and low- resolution images in early remote sensing research (Wharton, 1982; Letourneau et al., 2012). Improvements on determining the optimal neighborhood were later investigated, including (van der Kwast et al., 2011). It is straightforward to define neighborhoods in such grid-based methods, and these methods are widely used for pixel-based image analysis. The limitation is that such a regular neighborhood does not correspond to a real neighborhood on the ground, particular in VHR images. For those, an object-based image analysis is preferred (Blaschke, 2010). Graph-based methods facilitate object-based land use classification. A graph-based method starts from the construction of a planar graph, in which the contextual information can be calculated within a sub-graph. Interesting work on using graph-based methods for land use classification is given by Barnsley and Barr (1997); Comber et al. (2012); Walde et al. (2014). Such a graph can represent the spatial relations between land cover objects. The difficulty, however, is to effectively partition a graph into meaningful sub-graphs, cor-

responding to homogenous land use units. In recent years, computer vision techniques have been increasingly used to extract land cover and land use information from remote sensing images. Yang and Newsam (2010) conducted a scene-based land use classification using bag-of-visual-words, and published the UC Merced (UCM) land use dataset, which is now frequently used (Cheriyadat, 2014; Chen and Tian, 2014). Currently, interesting studies have used deep learning methods for scene-based land use classification (Castelluccio et al., 2015; Jean et al., 2016). These methods, particularly based upon convolutional neural networks (CNNs) (Krizhevsky et al., 2012), performed better than bag-of-visual-words. The last category for urban land use classification is vector-based methods. These methods rely on existing boundary data, such as road networks and parcels, to define homogenous land use units (Hu and Wang, 2013; Alahmadi et al., 2015; Li et al., 2016b). Various types of spatial information can therefore be derived. One concern of using such method is that existing vector data for defining land use units might not be available for some areas. In the following sections, we will further explore the role of a CNN in urban land use classification.

This paper aims to classify urban land use from VHR stereo satellite images, and particularly explores the use of open source data for improving classification accuracy. We will use a multilevel Bayesian framework for a vector-based classification of urban land use (Li et al., 2016b), starting with the classification of urban land cover, proceeding to the characterization of building types, and resulting in the classification of urban land use. The novelty of this paper is that

- A method is explored using VHR stereo images for urban land use classification.
- A Bayesian method is investigated to incorporate open source data for improving urban land use classification.
- A pretrained CNN is applied with fine tuning on the UCM dataset for extracting additional urban land use information.

The remainder of this paper is organized as follows: Section 5.2 describes the study area and data, and Section 5.3 illustrates the proposed urban land use classification from VHR stereo images. Section 5.4 presents experimental results and related analysis, followed by discussion in Section 5.5 and conclusions of this research in Section 5.6.

5.2 Study area and data

The study area is located in Oklahoma City, Oklahoma, the United States. Oklahoma City is the capital and the largest city of the state of Oklahoma. The total population was estimated to be about 631 thousand in the summer of 2015 with a total of 1.4 million in the entire Oklahoma City metro. Oklahoma City is representative of a typical American city which is not limited in its growth by topographic or water features.

A subset of GeoEye imagery in stereo mode was obtained, i.e. forward- and backward-view images, recorded on 17 October 2012. Both images

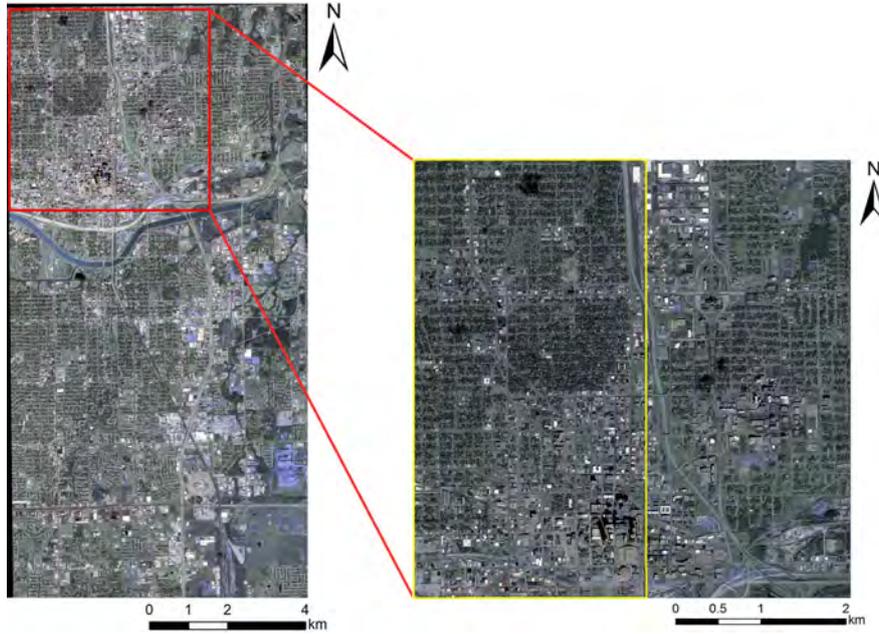


Figure 5.1: Overview of study area and datasets, depicted on a GeoEye-1 image in forward view of its stereo pair.

have four multispectral bands of 1.84 m and a panchromatic band of 0.5 m spatial resolution. Figure 5.1 provides an overview of the study area and images. The urban land use classification is mainly based upon the forward-view image, as it has a smaller off-nadir angle (5.48°) compared with the backward-view image (25.96°). This forward-view VHR image was fused using its panchromatic and multispectral channels using the Gram-Schmidt pan-sharpening method. The total study image (red box in Figure 5.1) has a size of 10363×10263 pixels. We selected a subset image (yellow box in Figure 5.1) of 10363×4813 pixels for convenient demonstration of the experiments. In addition, we used two publicly available datasets, i.e. Open Street Map (OSM) and the UCM land use dataset. The OSM of this study area was downloaded from: <https://www.geofabrik.de/data/download.html>. The UCM dataset is available at: <http://vision.ucmerced.edu/datasets/land-use.html>.

5.3 Methods

We applied a Bayesian framework for urban land use classification (Li et al., 2016b) (Figure 5.2). This framework covers the following steps: (1) generate a land cover classification from the forward-view VHR image; (2) characterize the spatial arrangement of the building types; (3) apply an urban land use classification using a Bayesian network classifier. In addition to this

5. Incorporating open source data for Bayesian classification of urban land use

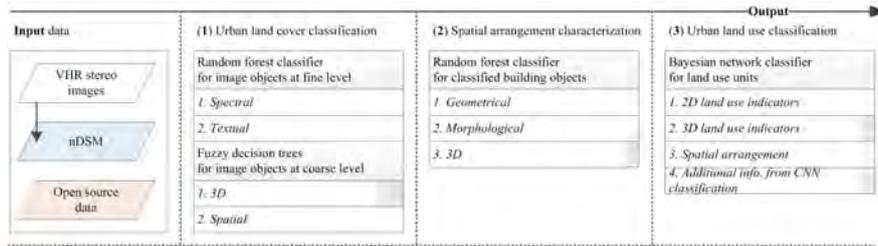


Figure 5.2: Framework of the Bayesian classification of urban land use from VHR stereo images incorporating open source data.

framework, this paper highlights the use of VHR stereo images, and the incorporation of open source data, namely an OSM and a UCM land use dataset, following Bayesian methodology. We used the road networks, including roads and railway, from the OSM to divide the VHR image into homogeneous land use units.

We used the VHR stereo images to generate a normalized Digital Surface Model (nDSM), from which height information was selected to improve the separability between elevated objects (i.e buildings) and non-elevated objects (other man-made objects). Fuzzy decision trees fine tuned by hierarchical Bayesian models conduct building extraction, where an OSM provides the prior building information. The UCM dataset is used to fine tune a pretrained convolutional neural network (CNN) (Castelluccio et al., 2015). This CNN is then applied to conduct an urban scene classification on our study image. The results of the scene classification are incorporated into the applied framework of land use classification, by Bayes' theorem.

5.3.1 nDSM generation from VHR stereo images

VHR stereo satellite images refer to those providing stereoscopic or tri-stereo pairs, e.g. from Geo-Eye, WorldView, and Pléiades satellites. By using photogrammetrical methods, these stereo images can be used to create a DSM. According to Aguilar et al. (2014), the vertical resolution of a DSM generated from stereo images can reach up to 2m, whereas local accuracy might vary. For image classification, we are interested in using an nDSM, which provides the height of objects above ground. An nDSM can be obtained by subtracting a Digital Terrain Model (DTM) from the corresponding DSM. In this paper, a DSM was obtained from VHR stereo images using the OrthoEngine in PCI Geomatica software, and its DTM was obtained by a filtering method, namely DSM2DTM provided by PCI. The final nDSM was generated by subtracting the DTM from the corresponding DSM. Although many other options for DSM and DTM generation can be found in the literature (Hirschmuller, 2008; Inglada and Christophe, 2009; Shean et al., 2016), we chose PCI because of its easy accessibility and satisfactory performance (Aguilar et al., 2014).

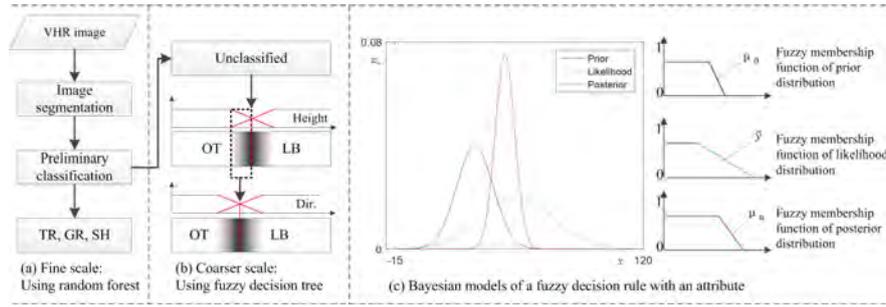


Figure 5.3: Workflow of urban land cover classification. (a) Preliminary land cover classification using a random forest at a fine scale. (b) Building extraction using fuzzy decision trees at a coarse scale. TR, GR, SH, LB, and OT refer to *trees*, *grass*, *shadow*, *buildings*, and *others*. Height and 'Dir.' are height and directional attributes for constructing fuzzy rules.

5.3.2 Incorporating OSM for urban land cover classification by Bayesian analysis

We distinguish five urban land cover classes: *trees*, *grass*, *shadow*, *buildings*, and *others*. A multiscale object-based method was used for urban land cover classification. We used the multiresolution segmentation method in eCognition software to create multiscale image segmentations (Benz et al., 2004). At a fine scale, image objects are preliminarily classified into *trees*, *grass*, *shadow* and unclassified (including *buildings* and *others*), using a random forest classifier (Figure 5.3a). Spectral and textural features of image objects are calculated for classification (Table 5A.1 in Appendix). At a coarse scale, the unclassified image objects are further classified into *buildings* and *others*, using fuzzy decision trees (Figure 5.3b). We constructed two fuzzy rules for decision trees, based upon height and directional features. The detailed description on the directional feature can be found in Li et al. (2015a, 2016a). This study uses hierarchical Bayesian models to determine the lower and upper bounds for these two fuzzy rules (van de Vlag and Stein, 2007) (Figure 5.3c). The next section describes in detail the construction of fuzzy decision trees based upon hierarchical Bayesian models with the incorporation of OSM data.

5.3.2.1 Fuzzy decision trees based upon hierarchical Bayesian models

For the adopted fuzzy decision trees, each fuzzy rule is formulated as a fuzzy membership function with respect to an attribute, i.e. height or directional attributes in this paper. We use linear membership functions for both rules that were also used in a previous study for a directional attribute (Li et al., 2015a). For each linear fuzzy membership function, its lower and upper bounds are determined in a Bayesian way (van de Vlag and Stein, 2007).

Let η be an unknown parameter, i.e. height or directional attributes, and $X = \{x_1, \dots, x_n\}$ be observed data. We aim to obtain the conditional

probability $p(\eta|X)$, based upon Bayes' theorem,

$$p(\eta|X) = \frac{p(\eta)p(X|\eta)}{p(X)} \propto p(\eta)p(X|\eta), \quad (5.1)$$

where $p(\eta|X)$, $p(\eta)$ and $p(X|\eta)$ refer to posterior, prior, and sampling distributions, respectively.

Let η follow a Normal distribution defined by a prior mean μ_0 and variance τ_0^2 . For normally distributed observations X , the sampling distribution is given by its sample mean \bar{x} and variance σ^2 . Then the posterior distribution is defined as (Gelman et al., 2013),

$$p(\eta|x_1, \dots, x_n) \sim \mathcal{N}(\eta|\mu_n, \tau_n^2), \quad (5.2)$$

where

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \quad (5.3)$$

$$\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}. \quad (5.4)$$

The prior probability of building height is derived from OSM and the generated nDSM by VHR stereo images. More specifically, building footprints can be obtained from an OSM. Based upon nDSM, the mean, and variance of the heights with respect to these building footprints are derived. In fact, we take those building footprints into account that have height values less than a threshold value ϵ_h , where we set ϵ_h equal to 3 m (referring to the height of single-story buildings) in this study. Our assumption is that an image object certainly belongs to a building, if its derived height value from nDSM is greater than ϵ_h , whereas low-rise building objects are uncertainly extracted relying on such nDSM. The fuzzy rule of the height attribute is fine tuned to deal with low-rise building image objects. For the directional attribute of a building, its prior probability is calculated based upon a set of randomly simulated values ranging from 0.5 to 1. The underlying assumption is that an image object is more likely to be a building, if its directional feature value is greater than 0.5 (Li et al., 2015a).

Finally, we take the lower and upper bounds of the 99.9% confidence interval of a posterior distribution as the corresponding lower and upper bounds of a linear membership function. The final land cover map is then obtained by overlaying the extracted *buildings* on the preliminary classified results.

5.3.3 Spatial arrangement characterization by building types

In our previous study, we demonstrated that the spatial arrangement information characterized by building types is effective for urban land use classification (Li et al., 2016b). We also indicated that building types are dependent

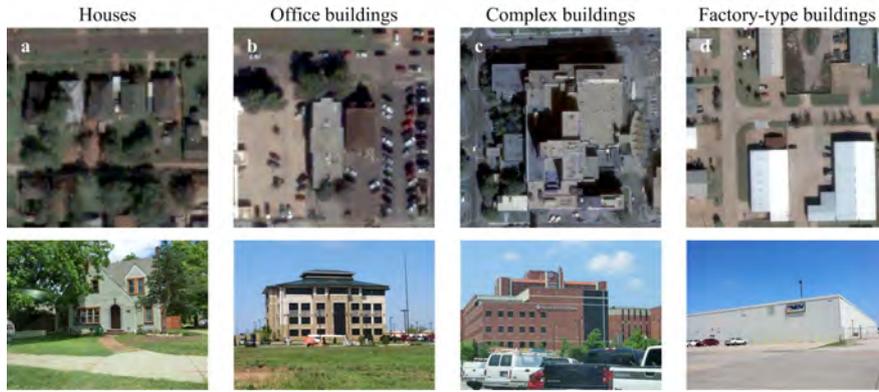


Figure 5.4: Building types visualized on VHR images (Top) and field survey photos (Bottom).

on urban structures, and might vary for different study areas. For this study area, we distinguish four building types: *houses*, *office buildings*, *complex buildings*, and *factory-type buildings* (Figure 5.4). These four building types were defined by both visually observing the building characteristics on the study images and a field survey of the study area. *Houses* mainly refer to residential buildings, characterized by small footprint and low height. *Houses* are usually surrounded by trees and grasses. *Office buildings* can be used for both residential and commercial purposes. We found that most of *office buildings* are used for commercial in our study area. They are usually larger and taller than *houses*. *Complex buildings* are usually used for multiple functions. In contrast to *office buildings*, *complex buildings* are characterized by a large footprint and irregular shapes. They can also be formed by several adjacent small buildings. *Factory-type buildings* are warehouses. They are usually larger than *houses*. When comparing with *office buildings* and *complex buildings*, *factory-type buildings* are more regular in shape and more homogeneous in spectral response.

These four building types are classified according to a set of geometrical, morphological, spatial and 3D features by a random forest classifier, see Table 5A.2 in the Appendix. Most of these features have been investigated in the literature (Voltersen et al., 2014; Li et al., 2016b). Within a land use unit, the spatial arrangement is then characterized by the distribution of classified building types.

5.3.4 Incorporating UCM dataset for urban land use classification with a CNN

5.3.4.1 CNN-based urban scene classification using UCM dataset

Development on urban scene classification methods has been greatly advanced by Yang and Newsam (2010), especially after the publication of the UCM dataset. We want to clarify that although Yang and Newsam (2010)

5. Incorporating open source data for Bayesian classification of urban land use

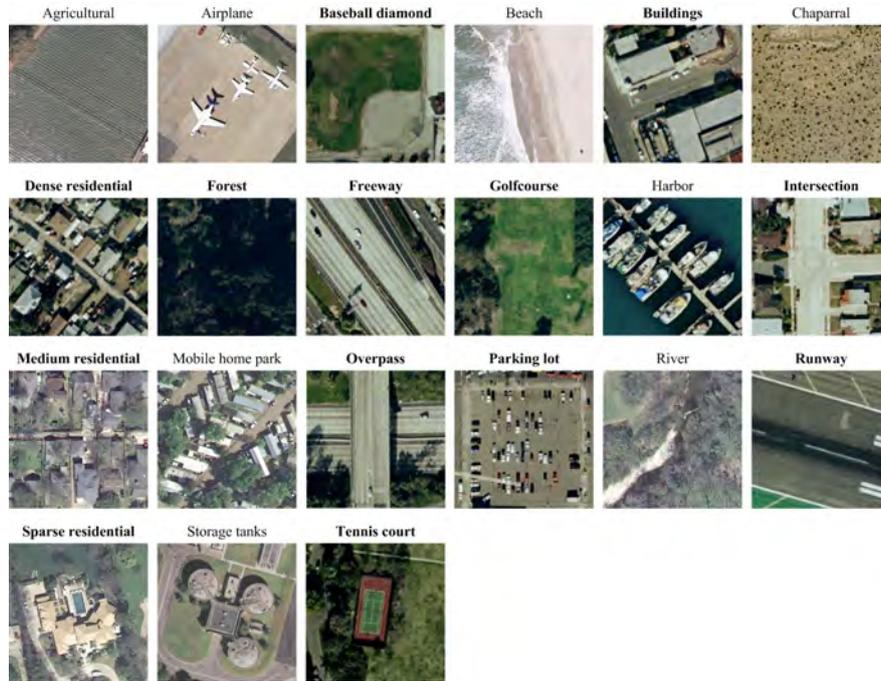


Figure 5.5: Overview of the UC Merced land use dataset, with 21 classes Yang and Newsam (2010). The classes in bold are used in this study.

also used the term "land use" in their research, their definition includes land cover classes (e.g. buildings), while our definition focuses solely on actual land use. We therefore call their "land use" classification as "scene" classification in this study. This UCM dataset was manually extracted from the USGS National Map Urban Area Imagery collection for 21 classes, with 100 images for each class and 256×256 pixels for each image (Figure 5.5). The scene content of these images is similar to that reflected by the study VHR image in a true-color composite. Hence, we assume that the UCM dataset could offer additional useful information for our urban land use classification. Furthermore, we selected 13 classes from the original 21 classes in the UCM dataset (see classes in bold of Figure 5.5) for this study, because these 13 classes can be found in our study area.

We use a CNN to conduct the urban scene classification on the study VHR images using the UCM dataset as a training dataset. This idea is inspired by Castelluccio et al. (2015), who explored the use of CNNs for urban scene classification on the UCM dataset. In this paper, we implemented a CNN-based scene classification on our study image in an easily accessible way. A detailed description on the theoretical parts of a CNN is out of the scope of this study but interested readers can refer to Castelluccio et al. (2015); Marmanis et al. (2016) for more details. In this study, we first normalized the VHR image (true-color composite) into $[0-255]$, and then divided the image into subsets with a size of 256×256

Table 5.1: Features used for land use classification.

| | | |
|---|---|---|
| 1. 2D Land use indicators (Voltersen et al., 2014; Li et al., 2016b) | | |
| <i>Geometrical</i> | Area | Perimeter |
| | Compactness | Shape Index |
| <i>Land cover related</i> | Coverage ratio of land cover | Density of land cover |
| <i>Landscape metrics</i> | Fractal dimension | Landscape shape index |
| | Shannon’s diversity index | |
| 2. 3D Land use indicators | | |
| Volume | Volume ratio of building and vegetation | Volume density of building and vegetation |
| 3. Spatial arrangement (by building types) (Li et al., 2016b) | | |
| $F_{SA} = [w_1 \cdot b_1, \dots, w_S \cdot b_S]^T$ | | |
| 4. Additional information (from CNN-based scene classification) | | |
| $F_S = [s_1, \dots, s_n]^T$, where s_i refers to the coverage ratio of a classified urban scene in a land use unit, and $n = 13$. | | |

allowing for 50% overlapping coverage. We applied a popular pretrained CNN-Caffe (Jia et al., 2014) for scene classification, which is available at: <http://www.vlfeat.org/matconvnet/models/imagenet-caffe-alex.mat>. Moreover, the MatConvNet toolbox was used (Vedaldi and Lenc, 2015) for implementation. Setting the parameters for fine tuning a pretrained CNN was mainly done by trial-and-error.

5.3.4.2 Bayesian classification of urban land use

We distinguish six urban land use classes: *residential*, *commercial*, *industrial and warehouses*, *green space and entertainment*, and *transportation*. The adopted framework for urban land use classification uses a Bayesian network classifier, which integrates commonly used land use indicators and spatial arrangement information (Li et al., 2016b). Table 5.1 lists features for land use classification.

Let $F = \{F_1, \dots, F_M, F_{SA}\}$ be a set of variables with respect to land use indicators (i.e. F_M) and the spatial arrangement (i.e. F_{SA}), and C_i be the states of class C with $i = 1, \dots, K$ the number of land use classes. Based upon Bayes’ theorem, the land use class of an unlabeled land use unit is thus inferred by

$$p(C_i|F) = \frac{p(C_i)p(F|C_i)}{p(F)}, i = 1, \dots, K. \quad (5.5)$$

For each unlabeled land use unit, we assign this land use unit to a category with maximum a posteriori (MAP) likelihood, given the input evidence. We then obtain

$$C^* = \operatorname{argmax}_{C_i} p(C_i)p(F|C_i), i = 1, \dots, K, \quad (5.6)$$

where $p(F)$ is the same for all the possible states of C_i . The prior probability $p(C_i)$ is equally likely for all land use classes, and the conditional probability

5. Incorporating open source data for Bayesian classification of urban land use

of $p(F|C_i)$ is estimated by kernels because of the limited number of training samples (Lin et al., 2007; Cheng and Wang, 2010; Maji et al., 2013).

We use the results of the urban scene classification as additional evidence. Equation 5.6 is thus transformed into

$$C^* = \underset{C_i}{\operatorname{argmax}} p(C_i)p(F|C_i)p(F_S|C_i), i = 1, \dots, K, \quad (5.7)$$

where F_S refers to the variable associated with scene classification results obtained from CNN. In this paper, F_S refers to a feature vector with respect to the proportion of each scene class within in a land use unit. We consider the variable of F_S independent of variables F , and estimate the conditional probability $p(F_S|C_i)$ also using kernels.

5.3.5 Evaluation of urban land use extraction

Urban land cover classification is evaluated by a confusion matrix, in which an overall accuracy (OA) and a kappa coefficient (κ) are also computed. The ground truth is obtained by visually selecting object samples from the study VHR image, for which a stratified sampling strategy is used. The building type classification and land use classification are evaluated by confusion matrices.

We partitioned the selected land use samples into training and testing datasets, using various random generators, ten different times. We compare the proposed land use classification, with land use classifications using just one of the four types of information, i.e. 2D land use indicators, 3D land use indicators, spatial arrangement information, and additional information from the UCM dataset. Here we label the proposed land use classification as $BN(F_{1,2,3,4})$, where $F_{1,2,3,4}$ represents the incorporation of all these four types of information. Similarly, we evaluate four additional classifications according to the incorporated information, namely $BN(F_{1,2,3})$, $BN(F_{1,3})$, $BN(F_{1,2})$, $BN(F_1)$, and $BN(F_2)$.

5.4 Results

5.4.1 Incorporating OSM for urban land cover classification by Bayesian analysis

We conducted image segmentations at two levels, labelled as L60 and L120, using the multi-resolution segmentation algorithm in eCognition software. These labels correspond to Scale parameter 60 and 120, respectively, whereas other parameters were set as default (i.e. Shape = 0.1 and Compactness = 0.5). We implemented the preliminary urban land cover classification using a random forest at segmentation level L60. To train the random forest classifier, we selected 60 object samples for each land cover class from the subset of the study image. Samples for the land cover classes *trees*, *grass* and *shadow* were selected at level L60, whereas samples for the *buildings* and *others* classes were selected at level L120.

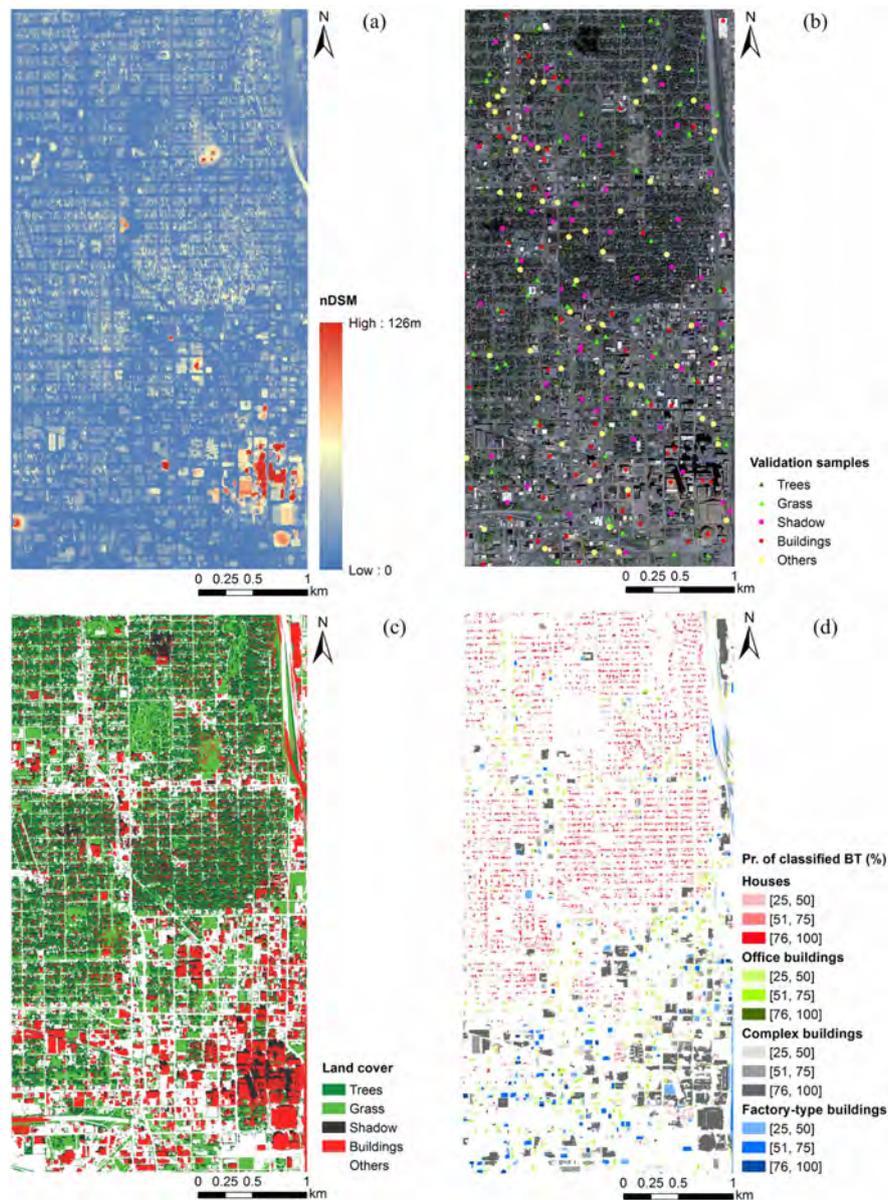


Figure 5.6: (a-d) are the derived nDSM, selected validation samples of urban land cover, classified urban land cover, and classified building types, with respect to the subset of the study image, respectively.

Fuzzy decision trees were used to separate building objects from other man-made objects at level L120. To derive parameters for the adopted fuzzy decision trees, we selected 100 building footprints from the OSM. The mean height of these buildings was obtained using the derived nDSM (Figure 5.6a).

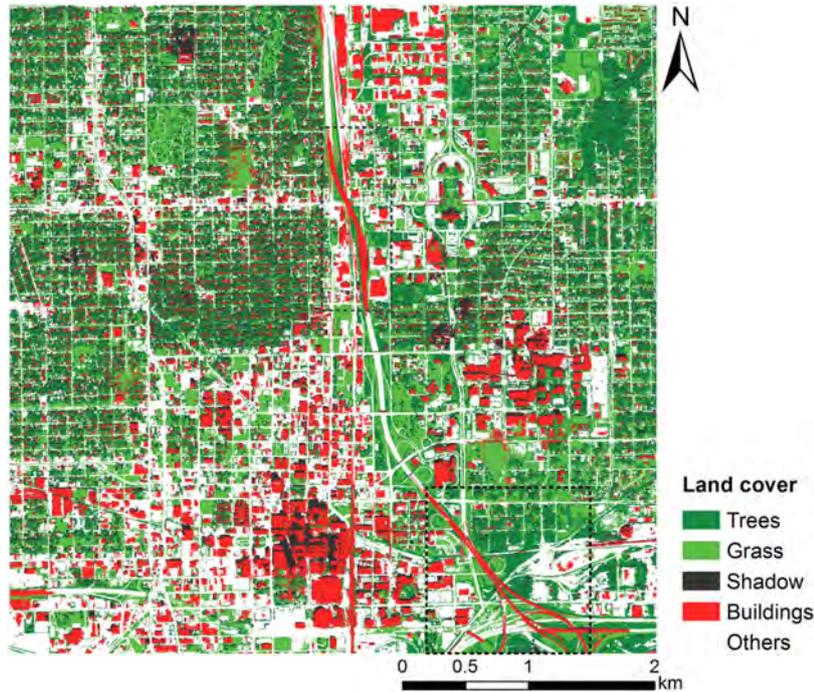


Figure 5.7: Urban land cover classified on the total study image.

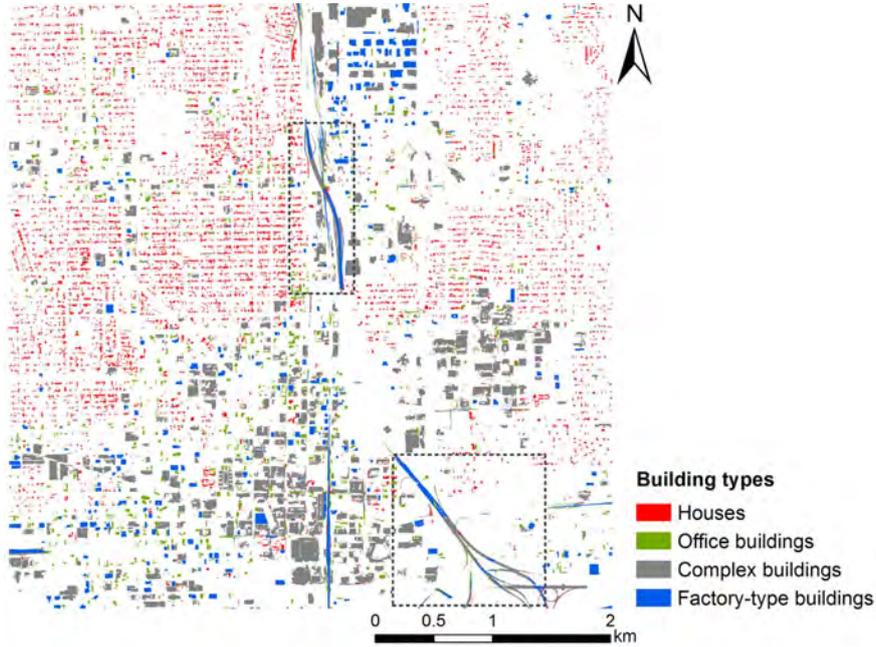
For those buildings having a mean height lower than ϵ_h , we computed the prior mean μ_0 and variance τ_0^2 with respect to the fuzzy rule of height attribute. The building samples, selected at object level L120, were used to compute parameters with respect to sampling distribution. More specifically, there were 15 objects with a height $< \epsilon_h$ used to derive the sample mean \bar{x} and variance σ^2 , where $n = 15$. After that, the posterior mean μ_n and variance τ_n^2 corresponding to the height rule were obtained. Similarly, the posterior mean and variance for the directional rule were computed. Different from its counterpart, i.e. directional rule, its prior mean and variance was computed from a set of simulated random numbers from 0.5 to 1. Hence, the lower and upper bounds for a fuzzy rule were chosen as the 99.9% confidence interval of the distribution defined by its posterior mean and variance.

Figure 5.6c shows the final urban land cover classification map by combining the results from both the random forest and fuzzy decision trees. Figure 5.7 shows the urban land cover classification for the total study image. 60 validation points for each land cover class were selected, based upon a stratified sampling strategy, for assessing the accuracy assessment of the land cover classification (Figure 5.6b). Here, five strata were generated by an unsupervised land cover classification on the subset image. Table 5.2 gives the confusion matrix of the urban land cover classification of the selected validation dataset. This table shows that the adopted urban land cover

Table 5.2: Confusion matrix of the urban land cover classification.

| Classified | Reference | | | | | UA |
|------------|-----------|-------|--------|----------|--------|-------|
| | Trees | Grass | Shadow | Building | Others | |
| Trees | 47 | 1 | 1 | 0 | 0 | 95.72 |
| Grass | 13 | 59 | 0 | 0 | 2 | 79.73 |
| Shadow | 0 | 0 | 58 | 1 | 0 | 98.31 |
| Building | 0 | 0 | 0 | 52 | 2 | 96.30 |
| Others | 0 | 0 | 1 | 7 | 56 | 87.50 |
| PA | 78.33 | 98.33 | 96.67 | 86.67 | 93.33 | % |

OA= 91%, $\kappa = 0.88$

**Figure 5.8:** Building types classified on the total study image.

classification had an overall accuracy (OA) of 91% and a kappa coefficient (κ) of 0.88. Only *trees* were classified with a producer accuracy (PA) below 85%, resulting into 13 *trees* samples being misclassified into *grass*. The classified *buildings* show a high PA. From the black boxes in Figure 5.7, we observed that some urban overpasses were misclassified into *buildings* because of height effects.

5.4.2 Spatial arrangement characterization by building types

The building type classification was first performed on the classified building objects using a random forest classifier. To train the classifier and validate its classification results, we selected 304 samples for the four building types

5. Incorporating open source data for Bayesian classification of urban land use

Table 5.3: Confusion matrix of the building type classification. Hou., O-B., C-B. and F-B. indicate building types *houses*, *office buildings*, *complex buildings* and *factory-type buildings*, respectively.

| Classified | Reference | | | | |
|---------------------------------------|-----------|-------|-------|-------|-------|
| | Hou. | O-B. | C-B. | F-B. | UA |
| Hou. | 52 | 1 | 0 | 0 | 98.11 |
| O-B. | 1 | 19 | 1 | 9 | 63.33 |
| C-B. | 0 | 0 | 29 | 2 | 93.55 |
| F-B. | 0 | 4 | 2 | 24 | 80.00 |
| PA | 98.11 | 79.12 | 90.63 | 68.57 | % |
| OA = 86%, $\kappa = 0.81$ | | | | | |
| Average: PA | 97.04 | 71.92 | 91.64 | 73.34 | |
| Average: OA = 85.49%, $\kappa = 0.80$ | | | | | |

defined for our study area, and randomly partitioned them into training and testing datasets. This partition was performed ten times by varying the random generator. The training dataset was represented by 40 samples for each building type. The remainder was assigned to the testing dataset. Figure 5.6d displays the classification results of the building types on the subset image, and visualizes the associated uncertainties in terms of classification probability as well. By visual inspection, we can see that all four building types were extracted. Table 5.3 gives the confusion matrix of the building type classification of the testing dataset, and average accuracies across the ten implementations. This table shows that the classification of building types was produced with an OA of 86% and a κ of 0.81, while the average OA was 85.49% and average κ was 0.80. More specifically, *houses* and *complex buildings* were classified with relatively higher accuracies than the other two selected building types. From this table, we also observed some evident misclassification between *office buildings* and *factory-type buildings*. There were 9 *factory-type buildings* misclassified into *office buildings*, by contrast, 4 *office buildings* were misclassified into *factory-type buildings*. The building type classification on the total study image is shown in Figure 5.8. The black boxes show that the misclassified buildings from overpasses were subsequently classified into *factory-type buildings* and *office buildings*.

5.4.3 Incorporating UCM dataset for urban land use classification with a CNN

First, we applied the pretrained CNN (Caffe) to our study VHR image. To fine tune the CNN and validate its performance on the UCM dataset, we randomly partitioned this dataset into a training (70) and testing (30) datasets. This pretrained CNN produced an OA of 92.31% on the testing dataset, see Figure 5.9 for the accuracy of each class (among the selected 13 scene classes). Next, the fine tuned CNN model was applied to the RGB image (in true color composite) of the study image. The scene classification result is provided in Figure 5.10. Although the CNN produced high classification accuracy on the UCM testing dataset, we observed that a large portion of a *residential* area was misclassified into *tennis court* (white box in Figure 5.10).

Second, we applied the urban land use classification to our study VHR im-

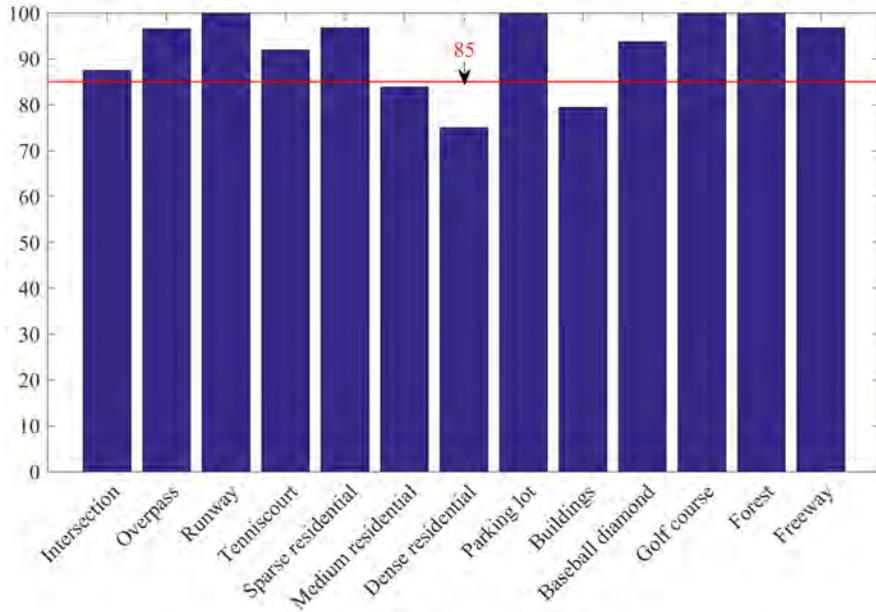


Figure 5.9: Accuracy of the scene classification using the pretrained CNN for the UCM testing dataset.

age, with an incorporation of the scene classification result (Equation 5.7). We selected 70 samples for each land use class, and then randomly partitioned them into a training (35) and testing (35) datasets. The samples were selected by visually interpreting the study image, with references of the corresponding OSM (i.e. its land use layer) and photos from our field survey. The training dataset was used to estimate conditional probabilities in Equation 5.7. The result of the classified urban land use is provided in Figure 5.11. We observed that the overpasses were classified into *transportation* with a high accuracy, although they were misclassified into *buildings* during land cover classification and subsequently into *factory-type buildings* and *complex buildings* during building type classification.

5.4.4 Evaluation of urban land use extraction

We first computed the confusion matrix on the testing dataset (Table 5.4). This table shows that the land use classification produced an OA of 88% and a κ of 0.85 on this testing dataset. Except for the *industrial* class, all land use classes were classified with competitive accuracies. *Transportation* had the highest classification, followed by *residential* which had a slight misclassification with *green space and entertainment*. The most evident misclassification existed between *industrial* and *commercial*, where 11 *industrial* land use units were misclassified into *commercial*, followed by the misclassification of 4 *commercial* units into *industrial*.

Second, we conducted a sensitivity analysis to compare the differences

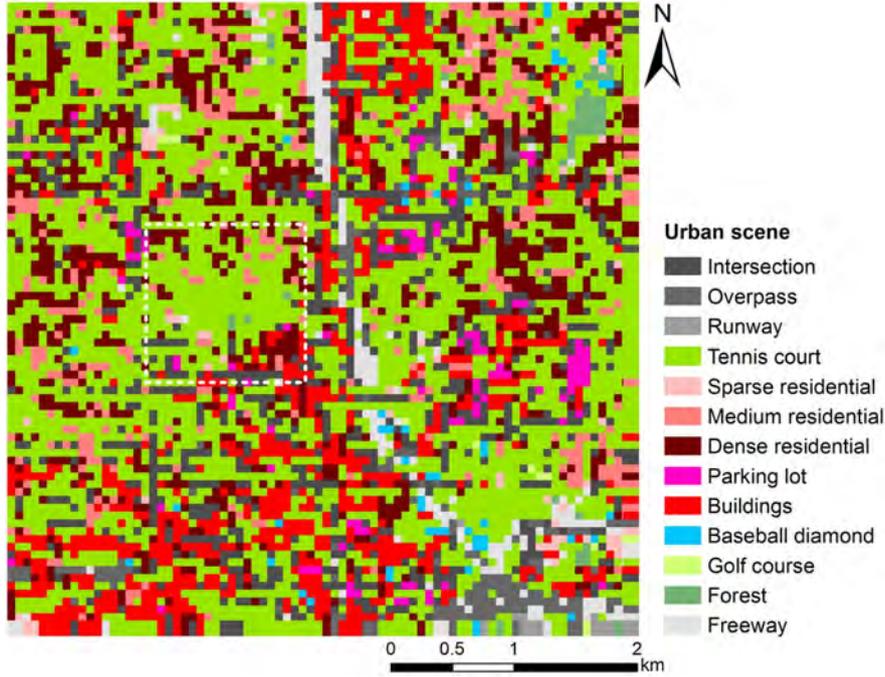


Figure 5.10: Urban scene classification on the study image using the pre-trained CNN fine tuned on UCM dataset.

Table 5.4: Confusion matrix of the urban land use extraction. Res., Com., Ind., G-E., and Tra. indicate the land use class *residential*, *commercial*, *industrial*, *green-entertainment* and *transportation*, respectively.

| Classified | Reference | | | | | UA |
|------------|-----------|-------|-------|-------|------|-------|
| | Res. | Com. | Ind. | G-E. | Tra. | |
| Res. | 33 | 0 | 0 | 2 | 0 | 94.23 |
| Com. | 0 | 31 | 11 | 1 | 0 | 72.01 |
| Ind. | 0 | 4 | 24 | 0 | 0 | 85.71 |
| G-E. | 2 | 0 | 0 | 31 | 0 | 93.94 |
| Tra. | 0 | 0 | 0 | 1 | 35 | 97.22 |
| PA | 94.29 | 85.57 | 68.57 | 88.57 | 100 | % |

OA = 88 %, $\kappa = 0.85$

between the proposed land use classification and those using less input information. Figure 5.12 provides the distributions of OA and κ across ten implementations for different land use classification. In general, we found lower classification accuracies if less input information was used. We found the highest accuracy when the information from the UCM dataset was incorporated. Moreover, VHR stereo images provide 3D information, which improved the classification performance compared to that using 2D information alone. We found remarkable differences between the collection $\{ BN(F_{1,2,3,4}), BN(F_{1,2,3}), BN(F_{1,3}) \}$ which used spatial arrangement information, and

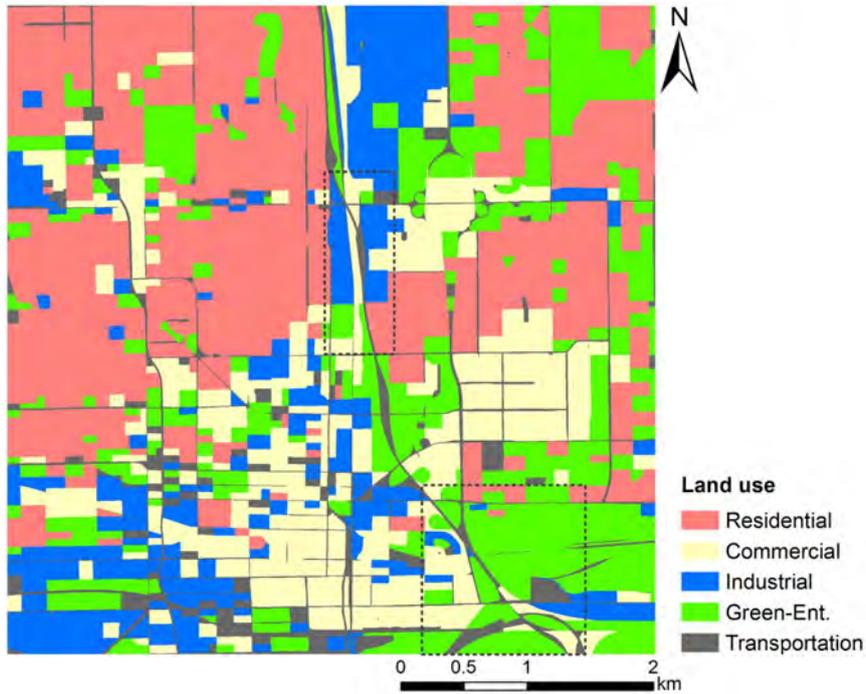


Figure 5.11: Urban land use classification on the study image.

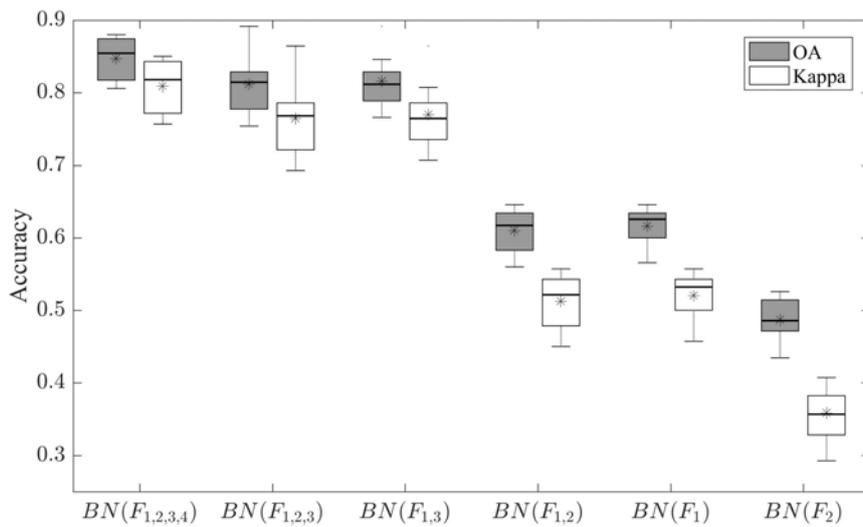


Figure 5.12: Distributions of OA and κ across ten various implementations for different land use classifications.

the collection $\{ BN(F_1), BN(F_1), BN(F_{1,2}) \}$ which did not use spatial arrangement information. Here the four types of information $F_{1,2,3,4}$ refer to

2D land use indicators, 3D land use indicators, spatial arrangement information, and additional information from the UCM dataset, respectively.

5.5 Discussion

In this study, we investigated urban land use classification from VHR stereo images, and explored the incorporation of open source data. Our first finding is that the adopted framework of urban land use classification is applicable to the study area in Oklahoma. This strengthens the usability of the framework developed in Li et al. (2016b). The second finding is that the 3D information derived from VHR stereo images not only improves urban land cover classification as illustrated in the literature (Longbotham et al., 2012), but also improves urban land use classification as illustrated in this study (Figure 5.12). The third finding is that the incorporation of open source data further improves urban land use classification. Furthermore, this paper demonstrates an urban land use classification combining the strengths of a knowledge-driven method, i.e. the Bayesian network classifier in Li et al. (2016b), and a data-driven method, i.e. the CNN in Castelluccio et al. (2015). To the best of our knowledge, no existing research has been carried out before to explore such combination for urban land use classification.

Hierarchical Bayesian models act as a tool to optimize the parameters of a fuzzy membership function by incorporating prior knowledge and collected samples. They have previously been investigated in the paper for a beach nourishment application (van de Vlag and Stein, 2007). This paper applied them to construct fuzzy rules for building extraction by incorporating knowledge from open source data, and produced satisfactory accuracies for classes *buildings* and *others*. Since this paper only uses height and directional features for building extraction, we found that some overpasses were misclassified into *buildings* (Figure 5.7). This misclassification in land cover however did not cause an evident impact on the subsequent land use classification (Figure 5.11). Still, as a previous study concluded (Li et al., 2016a), improvements in the land cover classification could aid in the improvement of the land use classification relying on it.

To transfer the adopted framework of land use classification to this study, we defined four building types and five land use classes for the Oklahoma study area. Figure 5.12 shows that the spatial arrangement information contributes the most to the land use classification accuracy within the collection of $\{BN(F_{1,2,3}), BN(F_{1,2}), BN(F_{1,3}), BN(F_1), BN(F_2)\}$. This result confirms the transferability of the adopted framework to a different study area. Besides, the use of 3D land use indicators increased the land use classification accuracy, when comparing $BN(F_{1,2,3})$ and $BN(F_{1,2})$ in Figure 5.12. This result demonstrates the advantage of using VHR stereo images for urban land use classification. In this paper, we also found evident misclassifications between *office buildings* and *factory-type buildings* (Table 5.3), and between *commercial* and *industrial buildings* (Table 5.4). The misclassification of land use might be proportionately affected by the misclassification of building types. This assumption could inspire future work to analyze such effects.

CNNs have greatly contributed to image categorization in computer vision. The remote sensing literature presents a number of interesting studies on the application of them for semantic information extraction from remote sensing images (Castelluccio et al., 2015; Jean et al., 2016). Usually, training a CNN from scratch is time-consuming, and it is impossible when a low number of training samples is available. In this context, the concept of transfer learning is useful to fine tune a pretrained CNN for a specific application. For example, Figure 5.9 shows that the classification accuracies of the selected 13 scene classes were highly competitive. In contrast, we found that the obtained urban scene result (Figure 5.10) hardly satisfies a practical application that requires clear land use boundaries. Furthermore, a large portion of *residential* area was misclassified into tennis court. When embedding the urban scene result (Figure 5.10) as an additional information source into the Bayesian framework of land use classification (Equation 5.7), however, we found that classification achieved the highest accuracy.

5.6 Conclusions

This study investigated a Bayesian classification of urban land use from VHR stereo images in particular with the incorporation of open source data, i.e. Open Street Map (OSM) and UC Merced (UCM) land use dataset. We applied a multilevel framework to the land use classification on a pair of GeoEye stereo images over Oklahoma City. Experimental results allow us conclude that: (1) the adopted land use framework is transferable to other study areas; (2) the use of VHR stereo images improves the urban land use classification compared with using a monoscopic VHR image alone; (3) the OSM data provide prior building information to construct fuzzy rules for building extraction, where hierarchical Bayesian models combine the prior information and observation evidence; (4) the scene classification of the VHR image provides additional evidence for land use classification, where a pretrained CNN with fine tuning on the UCM dataset serves as an effective scene classifier.

Appendix

Table 5A.1: Features used for land cover classification.

| 1. Preliminary classification (at a fine scale) | | | |
|--|------------------|-----------------|---------------|
| <i>Spectral</i> | Mean | StdDev | Brightness |
| | Max_Diff | NDVI | |
| <i>Textual (GLCM)</i> | Homogeneity | Contrast | Dissimilarity |
| | Entropy | Ang. 2nd moment | Mean |
| | StdDev | Correlation | |
| 2. Buildings extraction (at a coarse scale) | | | |
| <i>3D</i> | Height mean | | |
| <i>Spatial</i> | Directional mean | | |

Table 5A.2: Features used for building type classification.

| | | | |
|----------------------|--|-----------------|--------------|
| <i>Geometrical</i> | Area | Perimeter | Length/Width |
| | Border Index | Compactness | Shape Index |
| | Roundness | Rectangular Fit | Elliptic Fit |
| <i>Morphological</i> | Mean of morphological profiles | | |
| <i>3D</i> | Height mean | Volume | |
| <i>Contextual</i> | Coverage ratio of <i>vegetation</i> * and <i>others</i> within an image neighborhood** | | |

* *vegetation* refers to both *trees* and *grass*.

** An image neighborhood is defined as a square window that dilates the convex hull of a building object with 10 m.

Synthesis

6

6.1 Research findings and conclusions

This dissertation aims at developing methods to extract land use information from very high resolution (VHR) remote sensing images in urban areas. I conducted the urban land use extraction through analysis of the land cover derived from VHR images. This section presents the research findings and conclusions with respect to each research objective as described in section 1.7.

► **Objective 1: Develop a method to extract land cover over complex urban environments from VHR remote sensing images.**

Object-based image analysis methods have been the methods of choice used for urban land cover classification using VHR images. Two major challenges however are faced when implementing an object-based land cover classification: obtaining the optimal image segmentation from a VHR image and distinguishing buildings from other man-made objects. A binary partition tree (BPT) was used to build a hierarchical representation of an image, resulting into a series of image segmentations at multiple scales. An unsupervised evaluation of an image segmentation, based upon energy measures, was defined to choose the optimal segmentation from a BPT. The directional relationship between building and shadow objects was quantified by a fuzzy landscape to improve the separability between building and other man-made objects. Land cover classification was performed at two segmentation levels: the fine segmentation level was used to distinguish natural land cover classes, and the (coarser) optimal segmentation level was for building extraction. In addition, a set of geometrical error indices were modified to facilitate object-based accuracy assessment. Results showed that the proposed classification performed better than two other classifications using maximum likelihood and support vector machine, in terms of both pixel-based and object-based accuracy assessments.

This study concluded that the integration between BPTs and energy measures provided a means for unsupervised image segmentation, and the directional information could improve building extraction in urban areas. By doing so, the proposed method offered a potential solution for urban applications in need of obtaining land cover information from VHR images. Particularly, the quantified directional information was also used in Chapters 3, 4 and 5 for improving the separability of buildings in the Wuhan (China), Enschede (the Netherlands) and Oklahoma (United States) study areas. Despite these strengths, weakness of the proposed method can also be seen, which could lead to future research and improvement. These include the difficulty of using BPTs to decide a meaningful representation or a segmentation at a specific scale. In this study, the image segmentation obtained based upon the unsupervised evaluation was relatively optimal, but not perfect for all ground objects. Besides, the directional information relies on building

shadow, which may not be detectable in some areas. On the other hand, the shadow cast by trees and bridges can influence the accuracy of building extraction.

► **Objective 2: Develop a method to extract urban road from VHR remote sensing images.**

Road extraction from VHR images is important for many applications. For urban land use extraction, the roads detected from VHR images can help to obtain land use units, when land use boundaries or existing road data are not available. Since BPTs have potential for hierarchical image representation, it was further used to extract urban roads from VHR images. The road region of interest, preliminarily obtained from a VHR image by thresholding, was represented by a BPT. It reduced the effects of vegetation, shadow and buildings. Geometrical and structural features were considered within a region model and fused using fuzzy logic. The structural features were derived based upon orientation histograms and morphological profiles. Meaningful road regions were automatically extracted based upon two uncertainty related measures. Results showed that this method was able to group adjacent small segments that had high spectral heterogeneity and low road-like geometrical properties.

This study concluded that the proposed method could provide an alternative means for road extraction over dense urban areas from VHR images, including varying densities of roads, buildings adjacent to roads and partial covering of roads by vehicles and vegetation. Second, improvement in obtaining the road region of interest may improve the final road extraction and vice versa. Moreover, the BPT constructed in this study was application driven. This differed from a more common data driven BPT construction, and from the one built in Objective 1. An application driven BPT is more suited for a specific extraction task like road extraction, because it uses a specific region model and a merging criterion. In addition, an application driven BPT can make use of expert knowledge for BPT construction. The structural feature was derived from morphological profiles created by a series of path opening operations. This feature was also used in Chapters 4 and 5 for characterizing building types.

► **Objective 3: Develop a method to infer land use from the land cover derived from VHR remote sensing images in urban areas.**

A probabilistic method for urban land use extraction was developed using a Bayesian network. The spatial arrangement of land cover features was characterized. This characterization was done by quantifying the distribution of building types within a land use unit. Building types were classified from building objects based upon a set of geometrical, morphological, and contextual features. Land use indicators, primarily based upon the coverage ratio and density of land cover features, were combined with spatial arrangement information in this method, to in-

fer land use. Urban land use extraction was performed on a Pléiades image in the Wuhan study area, for which land use classes were generalized from a detailed classification system (of 45 land use classes) provided by local department. Results showed that integrating the spatial arrangement significantly improved the accuracy of urban land use extraction as compared to using land use indicators alone. The Bayesian network classifier produced results comparable to those produced by support vector machine and random forest classifiers.

This study concluded that the proposed characterization of spatial arrangement and Bayesian network integration was effective for urban land use extraction from VHR images. Moreover, the characterized spatial arrangement was able to provide an explicit interpretation for urban land use classes, compared to those using low-level image features such as SIFT to characterize the spatial information of an image. The proposed framework was transferable to other study areas, as it was also applied to the Oklahoma study area in Chapter 5. For a different urban area, we need to modify the definition of building types and land use classes based upon its urban structure and land use system (e.g. local official land use classification system), yet the proposed framework remains the same. The strength of the Bayesian network for modeling urban land use was that it was able to fuse different types of knowledge and handle associated uncertainties. Its difficulty is, however, to build a structure (i.e. a directed acyclic graph) and estimate its parameters. In this study, the Bayesian network structure was built based upon expert knowledge. The spatial arrangement variable was assumed independent of land use indicators. Although this assumption produced acceptable results in this paper, future improvement could be done by considering a dependence relationship between these two types of variables.

► **Objective 4: Apply the method developed in objective 3 to a different study area using VHR stereo images.**

Recent satellites like GeoEye, Worldview and Pléiades can provide VHR images with multiple views, corresponding to VHR stereo or tri-stereo images. These images can provide additional information, such as height information, for land cover and land use extraction. Open source data has also been increasingly accessible to the public. This not only brings opportunities for land use extraction using multiple sources, but also poses challenges for existing methods to be more robust and extendable. In this study, the developed method from Objective 3 was applied to the Oklahoma study area for urban land use extraction using VHR stereo images acquired from GeoEye satellite. The incorporation of open source data, i.e. Open Street Map (OSM) and the UC Merced (UCM) land use dataset, was also investigated under this framework. The OSM data was incorporated in the process of land cover classification by Bayesian analysis. The UCM dataset was used to fine tune a pretrained convolutional neural network (CNN) for

performing a scene-based classification. The result of the scene classification using this CNN was incorporated into the applied framework by Bayes's theorem. Results showed that the urban land use classification using VHR stereo images performed better than that using a monoscopic VHR image, the integration of open source data improved the final urban land use classification.

This study first concluded that the adopted method of urban land classification is applicable to the study area in Oklahoma. This could strengthen the usability of the method developed in this dissertation. Second, the use of VHR stereo images improved the urban land use classification compared with using a monoscopic VHR image alone. Third, the incorporation of open source data further improved urban land use classification, for which Bayesian analysis acts as an effective tool. Furthermore, this study demonstrated an urban land use classification combining the strengths of a knowledge-driven method, i.e. the Bayesian network classifier, and a data-driven method, i.e. the CNN. To the best of my knowledge, no such research was done before.

6.2 Reflections

Urban land use information plays a significant role in many urban related applications. Particularly, at the local scale, urban land use maps are capable of providing agencies, planners and researchers with information of fine spatial details to understand the status of urban environments, to analyze the interactions between humans and the environment, and to develop solutions for sustainable urban development. A land use map that is specified on the basis of parcels or street blocks and that can distinguish residential from commercial and industrial use, allows to better estimate urban growth and population density, and assess the quality of life and social vulnerability. Moreover, a land use map at a fine scale is able to produce maps at coarse scales by means of map generalization, serving a wider range of applications. Uncertainty is inevitably involved in the process of producing a land use map by any means, and might impair the reliability of the product. Handling the associated uncertainty is, therefore, important for a land use extraction system, and in turn for increasing the extraction accuracy. The method that was developed in this dissertation provides a means to extract urban land use at the local level, while taking associated uncertainty into account.

VHR remote sensing images, acquired for example from GeoEye, Worldview and Pléiades satellites, offer opportunities to implement land use extraction at the local scale. It is interesting in this context that the second national land use survey conducted by the Chinese Ministry of Land and Resources, largely adopted VHR images, with the incorporation of field survey and other ancillary data, to extract urban land use maps with a 1:500 mapping scale for urban areas. This survey still extracted land use information mainly by human interpretation, but it has emphasized the significance of using VHR images to extract land use maps.

Most of current image analysis methods are transferred from a wider set of methods that are initially developed for classifying land cover types. VHR images directly record the spectral reflectance of the earth surface, whereas urban land use reflects human activities. Several methods of land cover classification, based upon the exploitation of spectral, textual and geometrical information, therefore fail to achieve a satisfactory land use extraction, particularly from VHR images in urban areas. Moreover, the methods developed for land use extraction using low- and medium-resolution remote sensing images, are perhaps inefficient for using VHR images if more detailed spatial arrangement information is not characterized and used. In this dissertation, I focused on developing and applying advanced image analysis methods to improve the extraction of urban land use from VHR images. As major considerations for the methods in this thesis, I used Bayesian networks and BPTs.

Bayesian networks, i.e. probabilistic graphical models, can serve as an effective means to integrate various types of knowledge, and to handle associated uncertainty. In this dissertation, I developed a method for land use extraction from VHR images using a Bayesian network. This method infers land use from land cover objects derived from a VHR image, and integrates knowledge both from land use indicators that have been used in existing studies and from the spatial arrangement information that was characterized by a novel means proposed in this dissertation. This method was tested at two different study areas in Wuhan, China and Oklahoma, United States. Results in these two areas proved the effectiveness of this method. In this sense, the developed method should also be applicable to more other study areas for land use extraction. Besides, this dissertation also contributes to a set of targeted image analysis methods for extracting information from VHR images, including methods for urban land cover classification and urban road extraction. These methods offer attractive options for extracting corresponding information using VHR images.

BPTs can be used for hierarchical image representation. The representation of earth objects like land cover objects and roads, however, is scale-dependent. BPTs are therefore suited for hierarchically representing VHR images. Ideally, the construction of a BPT should be application driven, because in this sense the adopted region model and region merging criteria of the BPT can be well tuned according to a specific application.

Apart from the methodologies, also new sources of data are available in the era of big urban data. Here I considered open source data like OSMs and the UCM land use dataset. Open source data are not always well calibrated to the targeted application, but the available dataset is big in size and free to use. It thus may provide useful information to the application. This brings not only the opportunities for land use extraction but also challenges to the existing image analysis methods. In this dissertation, a way of incorporating open source data for land use extraction was investigated by means of Bayesian analysis, as part of the developed method for land use extraction using Bayesian networks. Results showed that this method could be extended to incorporate additional information from the open source data. Due to the strengths of Bayesian networks, the extension capability should not

be limited to incorporation of analysis those two types of open source data, but also others.

Uncertainty involved in every step of urban land use extraction. In this dissertation, I focused on handling uncertainties associated with (1) image objects' representation, (2) features' quantification, (3) knowledge quantification, (4) classifiers, and (5) object-based accuracy assessment. Uncertainty handling was done by both fuzzy sets and probabilistic methods. In chapter 2, the directional relationship between building and shadow objects was quantified by fuzzy sets taking direction and distance related elements into account. Moreover, a set of modified geometric error indices were proposed for object-based accuracy assessment, taking both local and global distributions of the geometric errors of classified image objects into account. In chapter 3, road regions were represented by a BPT, for which the uncertainty related to the region model and merging criteria of the BPT was handled by fuzzy sets. Moreover, two uncertainty measures, namely possibility and necessity measures, were used for road extraction based upon a BPT. In chapter 4, the spatial arrangement of land use types was characterized based upon building types, taking their classified probabilities into account. Moreover, land use extraction was conducted by a Bayesian network classifier. Expert knowledge was incorporated into land use extraction based upon Bayesian methods. In chapter 5, open source data were also incorporated to the developed land use extraction method based upon Bayesian methods. The methods that I used for uncertainty handling in this dissertation, are also potentially suitable for dealing with the uncertainty associated with a wider range of land information extractions from remote sensing images.

Besides the strengths of the developed method for land use extraction, I am aware of two main weaknesses. The first is regarding uncertainty propagation. In chapter 4, a sensitivity analysis was conducted to analyze the effect of land cover classification on the final land use extraction. Results showed that the accuracy of land cover classification proportionally affects the accuracy of land use extraction. That means that land use extraction is dependent on land cover classification, because the land use is inferred from the land cover derived using VHR images. The uncertainty propagation issue, however, is out of the scope of this dissertation. Addressing this issue could lead to an improvement of the developed land use extraction as presented in this dissertation. The other weakness is that the developed method needs users to specify building types and land use classes. This is to some degree reasonable, because building types and land use classes vary between different study areas. In this dissertation, I applied the developed land use extraction to different study areas in Wuhan City and Oklahoma City respectively. These two study areas are respectively typical for a moderate, booming city in China and a mid-western state capital in the US. Therefore, the adopted building types and land use classes could serve as a good reference to other study areas in the need of obtaining land use information from VHR images.

6.3 Recommendations

This dissertation has investigated the use of VHR remote sensing images for land use extraction in urban areas, and presented a set of image analysis methods to extract information from VHR images. Many aspects of the methods have not yet been fully addressed, and further investigations are needed. The following recommendations are made:

1. BPTs constructed based upon non-parametric methods were investigated in Chapters 2 and 3 for image segmentation and road extraction. They can also be constructed by parametric methods to represent region models and region merging criteria. A possible practice is to build a probabilistic BPT.
2. Directional relationship between building and shadow was effectively used for building extraction in this dissertation. It is however affected by the extracted shadow, which might be confused with water and dark roofs. This could decrease the accuracy of building extraction. Improvements in shadow extraction are needed to improve building extraction using such directional relationship.
3. Research on the structures and parameter estimation methods of Bayesian networks can further improve the developed method for land use extraction. For example, improvements can be done by adding the dependence between variables of land use indicators and spatial arrangement. Furthermore, more prior information on land use classes could be exploited.
4. The potential of using VHR stereo images should be further exploited for information extraction. VHR stereo images sometimes are acquired with poor quality, for example when the off-nadir angle is big. This can influence land cover classification based upon such images, and thus can decrease the extraction accuracy for land use.
5. Research can be conducted to investigate the potential of the developed land use extraction method for monitoring land use changes using multi-temporal VHR images. The straightforward strategy is to use a post-classification method. Moreover, the land use map extracted from an old VHR image, can be used as prior information for extracting land use from a new image.
6. Finally, converting the developed methods in this dissertation into an automatic extraction system is considered necessary in order to save efforts in need of producing land use maps. To do so, more different study areas and VHR images from various platforms should be investigated.

Bibliography

- Adeline, K., Chen, M., Briottet, X., Pang, S., Paparoditis, N., 2013. Shadow detection in very high spatial resolution aerial images: A comparative study. *ISPRS Journal of Photogrammetry and Remote Sensing* 80, 21–38.
- Aguilar, M., del Mar Saldaña, M., Aguilar, F. J., 2014. Generation and quality assessment of stereo-extracted DSM from GeoEye-1 and WorldView-2 imagery. *IEEE Transactions on Geoscience and Remote Sensing* 52 (2), 1259–1271.
- Akçay, H., Aksoy, S., 2010. Building detection using directional spatial constraints. In: 2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Honolulu, Hawaii, USA, 25-30 July, pp. 1932–1935.
- Alahmadi, M., Atkinson, P., Martin, D., 2015. Fine spatial resolution residential land-use data for small-area population mapping: a case study in Riyadh, Saudi Arabia. *International Journal of Remote Sensing* 36, 4315–4331.
- Alonso-González, A., López-Martínez, C., Salembier, P., 2014. PolSAR time series processing with binary partition trees. *IEEE Transactions on Geoscience and Remote Sensing* 52 (6), 3553–3567.
- Alonso-Gonzalez, A., Valero, S., Chanussot, J., Lopez-Martinez, C., Salembier, P., 2013. Processing multidimensional SAR and hyperspectral images with binary partition tree. *Proceedings of the IEEE* 101 (3), 723–747.
- Anwar, S., 2014. Spatial point process modelling of land use and land cover (LULC) change. Ph.D. thesis, University of Twente, Enschede, The Netherlands.
- Arnfield, A. J., 2003. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *International Journal of Climatology* 23 (1), 1–26.
- ASPRS, ASCE, 1994. Glossary of the mapping sciences. American Society of Civil Engineers.
- Bajcsy, R., Tavakoli, M., 1976. Computer recognition of roads from satellite pictures. *IEEE Transactions on Systems, Man and Cybernetics* SMC-6 (9), 623 – 637.

- Baltsavias, E., 2004. Object extraction and revision by image analysis using existing geodata and knowledge: Current status and steps towards operational systems. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3-4), 129–151.
- Banzhaf, E., Hofer, R., 2008. Monitoring urban structure types as spatial indicators with CIR aerial photographs for a more effective urban environmental management. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 1 (2), 129–138.
- Banzhaf, E., Netzband, M., 2011. Monitoring urban land use changes with remote sensing techniques. In: Richter, M., Weiland, U. (Eds.), *Applied Urban Ecology: A Global Framework*. Wiley-Blackwell, pp. 18–32.
- Barnsley, M., Barr, S., 1997. Distinguishing urban land-use categories in fine spatial resolution land-cover data using a graph-based, structural pattern recognition system. *Computers, Environment and Urban Systems* 21 (3-4), 209–225.
- Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3-4), 239–258.
- Bishop, C. M., 2006. *Pattern recognition and machine learning*. Springer.
- Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (1), 2–16.
- Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Feitosa, R. Q., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic object-based image analysis towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* 87 (0), 180–191.
- Bloch, I., 1999. Fuzzy relative position between objects in image processing: A morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (7), 657–664.
- Bloch, I., 2005. Fuzzy spatial relationships for image processing and interpretation: a review. *Image and Vision Computing* 23 (2), 89–110.
- Bouziani, M., Goita, K., He, D.-C., 2010. Rule-based classification of a very high resolution image in an urban environment using multispectral segmentation guided by cartographic data. *IEEE Transactions on Geoscience and Remote Sensing* 48 (8), 3198–3211.
- Brown, K. M., Foody, G. M., Atkinson, P. M., 2009. Estimating per-pixel thematic uncertainty in remote sensing classifications. *International Journal of Remote Sensing* 30 (1), 209–229.
- Buck, A., Keller, J., Skubic, M., 2013. A memetic algorithm for matching spatial configurations with the histograms of forces. *IEEE Transactions on Evolutionary Computation* 17 (4), 588–604.
- Butenuth, M., Heipke, C., 2012. Network snakes: Graph-based object delineation with active contour models. *Machine Vision and Applications* 23 (1), 91–109.

- Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., 2015. Land use classification in remote sensing images by convolutional neural networks. <https://arxiv.org/abs/1508.00092>.
- Castilla, G., Hay, G., 2007. Uncertainties in land use data. *Hydrology and Earth System Sciences* 11 (6), 1857–1868.
- Chai, D., Forstner, W., Lafarge, F., 2013. Recovering line-networks in images by junction-point processes. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Oregon, Portland, June 25-27, pp. 1894–1901.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2 (3), 27:1–27:27.
- Chen, S., Tian, Y., 2014. Pyramid of spatial relations for scene-level land use classification. *IEEE Transactions on Geoscience and Remote Sensing* 53 (4), 1947–1957.
- Cheng, H., Wang, R., 2010. Semantic modeling of natural scenes based on contextual Bayesian networks. *Pattern Recognition* 43 (12), 4042–4054.
- Cheriyadat, A., 2014. Unsupervised feature learning for aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing* 52 (1), 439–451.
- Clinton, N., Holt, A., Scarborough, J., Yan, L., Gong, P., 2010. Accuracy assessment measures for object-based image segmentation goodness. *Photogrammetric Engineering & Remote Sensing* 76 (3), 289–299.
- Clode, S., Rottensteiner, F., Kootsookos, P., Zelniker, E., 2007. Detection and vectorization of roads from Lidar data. *Photogrammetric Engineering & Remote Sensing* 73 (5), 517–535.
- Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (5), 603–619.
- Comber, A., Brunson, C., Farmer, C., 2012. Community detection in spatial networks: Inferring land use from a planar graph of land cover objects. *International Journal of Applied Earth Observation and Geoinformation* 18 (1), 274–282.
- Comber, A., Fisher, P., Brown, A., 2008. Uncertainty, vagueness, and indiscernibility: The impact of spatial scale in relation to the landscape elements. In: Stein, A., Shi, W., Bijker, W. (Eds.), *Quality Aspects in Spatial Data Mining*. CRC Press, pp. 239–250.
- Comber, A., Wadsworth, R., Fisher, P., 2010. Reasoning methods for handling uncertain information in land cover mapping. In: Devillers, R., Jean-soulin, R. (Eds.), *Fundamentals of Spatial Data Quality*. ISTE, pp. 123–139.
- Congalton, R., Green, K., 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. CRC Press.

- Dalal, N., Triggs, B., 2005. Histograms of Oriented Gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005. Vol. I. pp. 886–893.
- Ebert, A., Kerle, N., Stein, A., 2009. Urban social vulnerability assessment with physical proxies and spatial metrics derived from air- and spaceborne imagery and GIS data. *Natural Hazards* 48 (2), 275–294.
- Eyton, J., 1993. Urban land use classification and modelling using cover-type frequencies. *Applied Geography* 13, 111–121.
- FAO/UNEP, 1999. *The Future of Our Land: Facing the Challenge*. United Nations Environment Programme.
- Fauvel, M., Arbelot, B., Benediktsson, J., Sheeren, D., Chanussot, J., 2013. Detection of hedges in a rural landscape using a local orientation feature: From linear opening to path opening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (1), 15–26.
- Fauvel, M., Chanussot, J., Benediktsson, J., 2006. Decision fusion for the classification of urban remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 44 (10), 2828–2838.
- Fisher, P., 1999. Models of uncertainty in spatial data. *Geographical information systems* 1, 191–205.
- Foody, G. M., 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering & Remote Sensing* 70 (5), 627–633.
- Foody, G. M., 2010. Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sensing of Environment* 114 (10), 2271–2285.
- Gamba, P., Dell’Acqua, F., Stasolla, M., Trianni, G., Lisini, G., 2011. Limits and challenges of optical very-high-spatial-resolution satellite remote sensing for urban applications. In: Yang, X. (Ed.), *Urban Remote Sensing*. John Wiley & Sons, Ltd, pp. 35–48.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin, D., 2013. *Bayesian Data Analysis, Third Edition*. Taylor & Francis.
- Gerke, M., Butenuth, M., Heipke, C., Willrich, F., 2004. Graph-supported verification of road databases. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3-4), 152–165.
- Ghamisi, P., Dalla, M., Benediktsson, J., 2015. A survey on spectral-spatial classification techniques based on attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing* 53 (5), 2335–2353.
- GIZ, 2012. *Land use planning: concept, tools and applications*. Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.
- Grote, A., Heipke, C., Rottensteiner, F., 2012. Road network extraction in suburban areas. *The Photogrammetric Record* 27 (137), 8–28.
- Gruen, A., Li, H., 1997. Linear feature extraction with 3-D LSB-Snakes. In: Gruen, A., Kuebler, O., Agouris, P. (Eds.), *Automatic Extraction of Man-Made Objects from Aerial and Space Images*. Birkhäuser, pp. 287–298.

- Guigues, L., Cocquerez, J. P., Le Men, H., 2006. Scale-sets image analysis. *International Journal of Computer Vision* 68 (3), 289–317.
- Hamedianfar, A., Shafri, H. Z. M., Mansor, S., Ahmad, N., 2014. Improving detailed rule-based feature extraction of urban areas from WorldView-2 image and lidar data. *International Journal of Remote Sensing* 35 (5), 1876–1899.
- Heiden, U., Heldens, W., Roessner, S., Segl, K., Esch, T., Mueller, A., 2012. Urban structure type characterization using hyperspectral remote sensing and height information. *Landscape and Urban Planning* 105 (4), 361–375.
- Heijmans, H., Buckley, M., Talbot, H., 2005. Path openings and closings. *Journal of Mathematical Imaging and Vision* 22 (3), 107–119.
- Herold, M., Liu, X., Clarke, K., 2003. Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering & Remote Sensing* 69 (9), 991–1001.
- Heuvelink, G., Burrough, P., Stein, A., 1989. Propagation of errors in spatial modelling with GIS. *International Journal of Geographical Information Systems* 3 (4), 303–322.
- Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (2), 328–341.
- Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2010. A practical guide to support vector classification. <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- Hu, J., Razdan, A., Femiani, J., Cui, M., Wonka, P., 2007. Road network extraction and intersection detection from aerial images by tracking road footprints. *IEEE Transactions on Geoscience and Remote Sensing* 45 (12), 4144–4157.
- Hu, S., Wang, L., 2013. Automated urban land-use classification with remote sensing. *International Journal of Remote Sensing* 34 (3), 790–803.
- Huang, C. L., Wang, C. J., 2006. A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with Applications* 31 (2), 231–240.
- Inglada, J., Christophe, E., 2009. The orfeo toolbox remote sensing image processing software. In: *International Geoscience and Remote Sensing Symposium (IGARSS)*. Vol. 4. Cape Town, South Africa, July 12-17, pp. IV733–IV736.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 790–794.
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding. In: *ACM International Conference on Multimedia*. pp. 1–4.

- Krishnapuram, R., Keller, J. M., Ma, Y., 1993. Quantitative analysis of properties and spatial relations of fuzzy image regions. *IEEE Transactions on Fuzzy Systems* 1 (3), 222–233.
- Krizhevsky, A., Sutskever, I., Hinton, G., 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* 2, 1097–1105.
- Kurtz, C., Passat, N., Gañarski, P., Puissant, A., 2012. Extraction of complex patterns from multiresolution remote sensing images: A hierarchical top-down methodology. *Pattern Recognition* 45 (2), 685–706.
- Lackner, M., Conway, T., 2008. Determining land-use information from land cover through an object-oriented classification of IKONOS imagery. *Canadian Journal of Remote Sensing* 34 (2), 77–92.
- Lacoste, C., Descombes, X., Zerubia, J., 2005. Point processes for unsupervised line network extraction in remote sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (10), 1568–1579.
- Lang, S., Tiede, G., 2007. Definiens developer [definiens developer]. *GIS-Business* (9), 34–37.
- Letourneau, A., Verburg, P., Stehfest, E., 2012. A land-use systems approach to represent land-use dynamics at continental and global scales. *Environmental Modelling & Software* 33, 61–79.
- Li, M., Bijker, W., Stein, A., 2015a. Use of binary partition tree and energy minimization for object-based classification of urban land cover. *ISPRS Journal of Photogrammetry and Remote Sensing* 102, 48–61.
- Li, M., Stein, A., Bijker, W., Zhan, Q., 2016a. Region-based urban road extraction from VHR satellite images using binary partition tree. *International Journal of Applied Earth Observation and Geoinformation* 44, 217–225.
- Li, M., Stein, A., Bijker, W., Zhan, Q., 2016b. Urban land use extraction from very high resolution remote sensing imagery using a Bayesian network. *ISPRS Journal of Photogrammetry and Remote Sensing* 122, 192–205.
- Li, W., Chen, C., Su, H., Du, Q., 2015b. Local binary patterns and extreme learning machine for hyperspectral imagery classification. *IEEE Transactions on Geoscience and Remote Sensing* 53 (7), 3681–3693.
- Lin, H.-T., Lin, C.-J., Weng, R. C., 2007. A note on platts probabilistic outputs for support vector machines. *Machine learning* 68 (3), 267–276.
- Liu, Z., Shen, L., Zhang, Z., 2011. Unsupervised image segmentation based on analysis of binary partition tree for salient object extraction. *Signal Processing* 91 (2), 290–299.
- Longbotham, N., Chaapel, C., Bleiler, L., Padwick, C., Emery, W. J., Pacifici, F., 2012. Very high resolution multiangle urban classification analysis. *IEEE Transactions on Geoscience and Remote Sensing* 50 (4), 1155–1170.
- Lowry, J. H., Lowry, M. B., 2014. Comparing spatial metrics that quantify urban form. *Computers, Environment and Urban Systems* 44, 59 – 67.

- Lucieer, A., Stein, A., 2002. Existential uncertainty of spatial objects segmented from satellite sensor imagery. *IEEE Transactions on Geoscience and Remote Sensing* 40 (11), 2518–2521.
- Maji, S., Berg, A., Malik, J., 2013. Efficient classification for additive kernel SVMs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (1), 66–77.
- Marmanis, D., Datcu, M., Esch, T., Stilla, U., 2016. Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geoscience and Remote Sensing Letters* 13 (1), 105–109.
- Matsakis, P., Sjahputera, O., Keller, J., Marjamaa, J., 2004. The use of force histograms for affine-invariant relative position description. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (1), 1–18.
- Matsakis, P., Wendling, L., 1999. A new way to represent the relative position between areal objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (7), 634–643.
- Mena, J., 2003. State of the art on automatic road extraction for GIS update: A novel classification. *Pattern Recognition Letters* 24 (16), 3037–3058.
- Mnih, V., Hinton, G., 2010. Learning to detect roads in high-resolution aerial images. *Lecture Notes in Computer Science* 6316 LNCS (PART 6), 210–223.
- Montoya-Zegarra, J., Wegner, J., Ladicky, L., Schindler, K., 2014. Mind the gap: Modeling local and global context in (road) networks. In: Jiang, X., Hornegger, J., Koch, R. (Eds.), *Pattern Recognition*. Vol. 8753 of *Lecture Notes in Computer Science*. Springer International Publishing, pp. 212–223.
- Movaghati, S., Moghaddamjoo, A., Tavakoli, A., 2010. Road extraction from satellite images using particle filtering and extended kalman filtering. *IEEE Transactions on Geoscience and Remote Sensing* 48 (7), 2807–2817.
- Müller, D., Munroe, D. K., 2014. Current and future challenges in land-use science. *Journal of Land Use Science* 9 (2), 133–142.
- Mumford, D., Shah, J., 1989. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics* 42 (5), 577–685.
- Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment* 115 (5), 1145–1161.
- Nielsen, T. D., Jensen, F. V., 2007. *Bayesian networks and decision graphs*. Springer Science & Business Media.
- Niemeyer, J., Rottensteiner, F., Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 87, 152–165.
- Novack, T., Kux, H., Feitosa, R. Q., Costa, G. A. O. P., 2014. A knowledge-based, transferable approach for block-based urban land-use classification. *International Journal of Remote Sensing* 35 (13), 4739–4757.

- OGC, 2012. OGC city geography markup language (CityGML) encoding standard. <http://www.opengeospatial.org/standards/citygml>.
- Ok, A., Senaras, C., Yuksel, B., 2013. Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing* 51 (3), 1701–1717.
- Otsu, N., 1979. Threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics SMC-9* (1), 62–66.
- Pacifici, F., Chini, M., Emery, W., 2009. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment* 113 (6), 1276–1292.
- Patino, J. E., Duque, J. C., 2013. A review of regional science applications of satellite remote sensing in urban settings. *Computers, Environment and Urban Systems* 37 (0), 1–17.
- Persello, C., Bruzzone, L., 2010. A novel protocol for accuracy assessment in classification of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing* 48 (3), 1232–1244.
- Pesaresi, M., Guo, H., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M., Kemper, T., Lu, L., Marin-Herrera, M. A., Ouzounis, G. K., Scavazzon, M., Soille, P., Syrris, V., Zanchetta, L., 2013. A global human settlement layer from optical HR/VHR RS data: Concept and first results. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (5), 2102–2131.
- Philipp-Foliguet, S., Guigues, L., 2008. Multi-scale criteria for the evaluation of image segmentation algorithms. *Journal of Multimedia* 3 (5), 42–56.
- Pozzi, F., Small, C., 2005. Analysis of urban land cover and population density in the United States. *Photogrammetric Engineering & Remote Sensing* 71 (6), 719–726.
- Ravanbakhsh, M., Heipke, C., Pakzad, K., 2008. Road junction extraction from high-resolution aerial imagery. *Photogrammetric Record* 23 (124), 405–423.
- Salembier, P., Garrido, L., 2000. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Transactions on Image Processing* 9 (4), 561–576.
- Sebari, I., He, D.-C., 2013. Automatic fuzzy object-based analysis of VHSR images for urban objects extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 79, 171–184.
- Serra, J., Kiran, B., 2013. Optima on hierarchies of partitions. In: Hendriks, C. L., Borgfors, G., Strand, R. (Eds.), *Mathematical Morphology and Its Applications to Signal and Image Processing*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 147–158.
- Seto, K. C., Fragkias, M., Güneralp, B., Reilly, M. K., 2011. A meta-analysis of global urban land expansion. *PLOS ONE* 6 (8), 1–9.

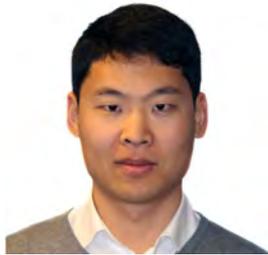
- Shean, D., Alexandrov, O., Moratto, Z., Smith, B., Joughin, I., Porter, C., Morin, P., 2016. An automated, open-source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 116, 101–117.
- Shi, W., 2009. *Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses*. CRC Press.
- Shi, W., Miao, Z., Wang, Q., Zhang, H., 2014. Spectral-spatial classification and shape features for urban road centerline extraction. *IEEE Geoscience and Remote Sensing Letters* 11 (4), 788–792.
- Shyu, C.-R., Klaric, M., Scott, G., Barb, A., Davis, C., Palaniappan, K., 2007. GeoIRIS: Geospatial information retrieval and indexing system mdash; content mining, semantics modeling, and complex queries. *IEEE Transactions on Geoscience and Remote Sensing* 45 (4), 839–852.
- Sigurdsson, E., Valero, S., Benediktsson, J., Chanussot, J., Talbot, H., Stefánsson, E., 2014. Automatic retinal vessel extraction based on directional mathematical morphology and fuzzy classification. *Pattern Recognition Letters* 47, 164–171.
- Soille, P., Pesaresi, M., 2002. Advances in mathematical morphology applied to geoscience and remote sensing. *IEEE Transactions on Geoscience and Remote Sensing* 40 (9), 2042–2055.
- Song, M., Civco, D., 2004. Road extraction using SVM and image segmentation. *Photogrammetric Engineering & Remote Sensing* 70 (12), 1365–1371.
- Stehman, S. V., Wickham, J. D., 2011. Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment. *Remote Sensing of Environment* 115 (12), 3044–3055.
- Stein, A., Hamm, N. A. S., Ye, Q., 2009. Handling uncertainties in image mining for remote sensing studies. *International Journal of Remote Sensing* 30 (20), 5365–5382.
- Stoica, R., Descombes, X., Zerubia, J., 2004. A Gibbs point process for road extraction from remotely sensed images. *International Journal of Computer Vision* 57 (2), 121–136.
- Tang, X., Song, S., Zhan, Y., 2014. A novel road extraction algorithm for high resolution remote sensing images. *Applied Mathematics and Information Sciences* 8 (3), 1435–1443.
- Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., Dech, S., 2012. Monitoring urbanization in mega cities from space. *Remote Sensing of Environment* 117 (0), 162–176.
- Teke, M., Başeski, E., Ok, A. O., Yüksel, B., Şenaras, C., 2011. Multi-spectral false color shadow detection. *Lecture Notes in Computer Science* 6952 LNCS, 109–119.
- Thakur, J., Singh, S., Ramanathan, A., Prasad, M., Gossel, W., 2012. *Geospatial Techniques for Managing Environmental Resources*. Springer.

- Tian, J., Chen, D. M., 2007. Optimization in multi-scale segmentation of high-resolution satellite images for artificial feature recognition. *International Journal of Remote Sensing* 28 (20), 4625–4644.
- Turetken, E., Benmansour, F., Andres, B., Pfister, H., Fua, P., 2013. Reconstructing loopy curvilinear structures using integer programming. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Oregon, Portland, June 25-27, pp. 1822–1829.
- Turetken, E., Benmansour, F., Fua, P., 2012. Automated reconstruction of tree structures using path classifiers and mixed integer programming. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. RI, USA, June 16-21, pp. 566–573.
- UN-Habitat, 2016. Urbanization and development: Emerging futures. <http://wcr.unhabitat.org/main-report/>.
- Ünsalan, C., Boyer, K. L., 2011. Review on land use classification. In: Ünsalan, C., Boyer, K. L. (Eds.), *Multispectral Satellite Image Understanding*. Springer London, pp. 49–64.
- Ünsalan, C., Sirmacek, B., 2012. Road network detection using probabilistic and graph theoretical methods. *IEEE Transactions on Geoscience and Remote Sensing* 50 (11 PART1), 4441–4453.
- Vaduva, C., Gavat, I., Datcu, M., 2013. Latent dirichlet allocation for spatial analysis of satellite images. *IEEE Transactions on Geoscience and Remote Sensing* 51 (5), 2770–2786.
- Valero, S., Chanussot, J., Benediktsson, J., Talbot, H., Waske, B., 2010. Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images. *Pattern Recognition Letters* 31 (10), 1120–1127.
- Valero, S., Salembier, P., Chanussot, J., 2013. Hyperspectral image representation and processing with binary partition trees. *IEEE Transactions on Image Processing* 22 (4), 1430–1443.
- Valero, S., Salembier, P., Chanussot, J., 2015. Object recognition in hyperspectral images using binary partition tree representation. *Pattern Recognition Letters* 56, 45–51.
- van de Vlag, D., Stein, A., 2007. Incorporating uncertainty via hierarchical classification using fuzzy decision trees. *IEEE Transactions on Geoscience and Remote Sensing* 45 (1), 237–245.
- van der Kwast, J., de Voorde, T. V., Canters, F., Uljee, I., Looy, S. V., Engelen, G., 2011. Inferring urban land use using the optimised spatial reclassification kernel. *Environmental Modelling & Software* 26 (11), 1279–1288.
- Vanegas, M., Bloch, I., Inglada, J., 2013. Alignment and parallelism for the description of high-resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 51 (6), 3542–3557.
- Vedaldi, A., Lenc, K., 2015. MatConvNet: Convolutional neural networks for MATLAB. In: *ACM International Conference on Multimedia*. pp. 1–55.

- Vilaplana, V., Marques, F., Salembier, P., 2008. Binary partition trees for object detection. *IEEE Transactions on Image Processing* 17 (11), 2201–2216.
- Voltersen, M., Berger, C., Hese, S., Schmallius, C., 2014. Object-based land cover mapping and comprehensive feature calculation for an automated derivation of urban structure types at block level. *Remote Sensing of Environment* 154, 192–201.
- Vosselman, G., Gunst, M., 1997. Updating road maps by contextual reasoning. In: Gruen, A., Kuebler, O., Agouris, P. (Eds.), *Automatic Extraction of Man-Made Objects from Aerial and Space Images*. Birkhäuser, pp. 267–276.
- Vosselman, G., Knecht, J., 1997. Road tracing by profile matching and kalman filtering. In: Gruen, A., Kuebler, O., Agouris, P. (Eds.), *Automatic Extraction of Man-Made Objects from Aerial and Space Images*. Birkhäuser, pp. 265–274.
- Walde, I., Hese, S., Berger, C., Schmallius, C., 2014. From land cover-graphs to urban structure types. *International Journal of Geographical Information Science* 28 (3), 584–609.
- Webpage, 2015. Open Street Map. <http://www.openstreetmap.nl/>.
- Wegner, J., Montoya-Zegarra, J., Schindler, K., 2013. A higher-order CRF model for road network extraction. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Oregon, Portland, June 25-27, pp. 1698–1705.
- Wharton, S. W., 1982. A contextual classification method for recognizing land use patterns in high resolution remotely sensed data. *Pattern Recognition* 15 (4), 317–324.
- Wickham, J. D., Stehman, S. V., Gass, L., Dewitz, J., Fry, J. A., Wade, T. G., 2013. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment* 130, 294–304.
- Wu, S.-S., Qiu, X., Usery, E. L., Wang, L., 2009a. Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use. *Annals of the Association of American Geographers* 99 (1), 76–98.
- Wu, S.-S., Xu, B., Wang, L., 2006. Urban land-use classification using variogram-based analysis with an aerial photograph. *Photogrammetric Engineering & Remote Sensing* 72, 813–822.
- Wu, Z., Yi, L., Zhang, G., 2009b. Uncertainty analysis of object location in multi-source remote sensing imagery classification. *International Journal of Remote Sensing* 30 (20), 5473–5487.
- Wurm, M., Schmitt, A., Taubenbock, H., 2016. Building types classification using shape-based features and linear discriminant functions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9 (5), 1901–1912.

- Yang, Y., Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification. In: 18th International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL). San Jose, United States, pp. 270–279.
- Zadeh, L. A., 1994. Fuzzy logic, neural networks, and soft computing. *Communications of the ACM* 37 (3), 77–84.
- Zhan, Q., Molenaar, M., Tempfli, K., 2002. Hierarchical image-object based structural analysis toward urban land use classification using high-resolution imagery and airborne lidar data. In: *Proceedings of the Remote Sensing of Urban Areas*. Istanbul, Turkey, pp. 251–258.
- Zhang, C., Murai, S., Baltsavias, E., 1999. Road network detection by mathematical morphology. In: *ISPRS Workshop 3D Geospatial Data Production: Meeting Application Requirements*. Paris, France, pp. 185–200.
- Zhang, X., Du, S., 2015. A linear dirichlet mixture model for decomposing scenes: Application to analyzing urban functional zonings. *Remote Sensing of Environment* 169, 37–49.
- Zhang, X., Du, S., Wang, Y. C., 2015. Semantic classification of heterogeneous urban scenes using intrascene feature similarity and interscene semantic dependency. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (5), 2005–2014.
- Zhao, B., Zhong, Y., Zhang, L., 2016. A spectral - structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 116, 73–85.
- Zhao, X., Stein, A., Chen, X., Zhang, X., 2011. Quantification of extensional uncertainty of segmented image objects by random sets. *IEEE Transactions on Geoscience and Remote Sensing* 49 (7), 2548–2557.

Biography



Mengmeng Li was born on 13 March, 1988 in Shandong, China. He obtained the BSc degree in Geographical Information Science at the Shandong University of Science and Technology in 2009, Qingdao, and the MSc degree in Cartography and Geographic Information System at the Fuzhou University in 2012, Fuzhou, China. In January 2013, he went to the Faculty of Geo-Information Science and Earth Observation (ITC) at the University of Twente for his PhD research, holding the scholarship from China Scholarship Council (CSC). During March to May 2016, he was an exchange PhD student at the University of Oklahoma, United

States. His PhD research focused on developing and applying image analysis methods to extract land use information from very high resolution (VHR) remote sensing image in urban areas. This thesis is the output of his PhD research.

Author's publication

ISI journal article under review

- **Li, M.**, de Beurs, K.M., Stein, A., Bijker, W. Incorporating open source data for Bayesian classification of urban land use from VHR stereo images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (major revision).

Published ISI journal articles

- **Li, M.**, Stein, A., Bijker, W., Zhan, Q., 2016. Urban land use extraction from Very High Resolution remote sensing imagery using a Bayesian network. *ISPRS Journal of Photogrammetry and Remote Sensing* 122, 192–205.
- **Li, M.**, Stein, A., Bijker, W., Zhan, Q., 2016. Region-based Urban Road Extraction from VHR Satellite Images Using Binary Partition Tree. *International Journal of Applied Earth Observation and Geoinformation* 44, 217–225.
- **Li, M.**, Bijker, W., Stein, A., 2015. Use of Binary Partition Tree and energy minimization for object-based classification of urban land cover. *ISPRS Journal of Photogrammetry and Remote Sensing* 102, 48–61.

Conference presentations

- **Li, M.**, Stein, A., Bijker, W., 2016. Urban land use extraction from very high resolution remote sensing imagery by Bayesian Network. *IG-ARSS 2016*, 10-15 July, Beijing, China.
- **Li, M.**, Stein, A., Bijker, W., de Beurs, K.M., 2016. Urban land cover and land use extraction from very high resolution remote sensing imagery. *US-IALE 2016*, 3-7 April, Asheville, United States.
- **Li, M.**, Stein, A., Bijker, W., 2015. Characterizing spatial arrangements for urban land use classification from Very High Resolution remote sensing images. *ILUS2015*, 11-13 November, Dresden, Germany.

ITC Dissertation List

http://www.itc.nl/Pub/research_programme/Research-review-and-output/PhD-Graduates