Automated Building Information Extraction and Evaluation from High-resolution Remotely Sensed Data

Chuiqing Zeng
The University of Western Ontario

Supervisor
Dr. Jinfei Wang
The University of Western Ontario

Graduate Program in Geography

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Chuiqing Zeng 2014

Follow this and additional works at: http://ir.lib.uwo.ca/etd
Part of the Geographic Information Sciences Commons, and the Remote Sensing Commons

Recommended Citation
Zeng, Chuiqing, 'Automated Building Information Extraction and Evaluation from High-resolution Remotely Sensed Data' (2014).
Electronic Thesis and Dissertation Repository. 2076.
http://ir.lib.uwo.ca/etd/2076

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca.
Automated Building Information Extraction and Evaluation from High-resolution Remotely Sensed Data

(Thesis format: Integrated Article)

by

Chuiqing Zeng

Graduate Program in Geography

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Chuiqing Zeng 2014
Abstract

The two-dimensional (2D) footprints and three-dimensional (3D) structures of buildings are of great importance to city planning, natural disaster management, and virtual environmental simulation. As traditional manual methodologies for collecting 2D and 3D building information are often both time consuming and costly, automated methods are required for efficient large area mapping. It is challenging to extract building information from remotely sensed data, considering the complex nature of urban environments and their associated intricate building structures.

Most 2D evaluation methods are focused on classification accuracy, while other dimensions of extraction accuracy are ignored. To assess 2D building extraction methods, a multi-criteria evaluation system has been designed. The proposed system consists of matched rate, shape similarity, and positional accuracy. Experimentation with four methods demonstrates that the proposed multi-criteria system is more comprehensive and effective, in comparison with traditional accuracy assessment metrics.

Building height is critical for building 3D structure extraction. As data sources for height estimation, digital surface models (DSMs) that are derived from stereo images using existing software typically provide low accuracy results in terms of rooftop elevations. Therefore, a new image matching method is proposed by adding building footprint maps as constraints. Validation demonstrates that the proposed matching method can estimate building rooftop elevation with one third of the error encountered when using current commercial software.

With an ideal input DSM, building height can be estimated by the elevation contrast inside and outside a building footprint. However, occlusions and shadows cause indistinct building edges in the DSMs generated from stereo images. Therefore, a “building-ground elevation difference model” (EDM) has been designed, which describes the trend of the elevation difference between a building and its neighbours, in order to find elevation
values at bare ground. Experiments using this novel approach report that estimated building height with 1.5m residual, which out-performs conventional filtering methods.

Finally, 3D buildings are digitally reconstructed and evaluated. Current 3D evaluation methods did not present the difference between 2D and 3D evaluation methods well; traditionally, wall accuracy is ignored. To address these problems, this thesis designs an evaluation system with three components: volume, surface, and point. As such, the resultant multi-criteria system provides an improved evaluation method for building reconstruction.

**Keywords**

building footprint, building boundary, building height, 3D structure, accuracy assessment, shape similarity, stereo image matching, high-resolution imagery, digital surface model, LiDAR, remote sensing, photogrammetry
Co-Authorship Statement

This thesis was prepared according to the integrated-article layout designed by the Faculty of Graduate Studies at Western University, London, Ontario, Canada. All the work stated in this thesis including methodology development, experimental testing, data analysis, modelling and writing draft manuscripts for publication was carried out by the author under the supervision of Dr. J. Wang. Versions of Chapters 2, 3, 4 and 5 have been published, accepted or submitted as co-authored peer reviewed journal papers. The co-authors can be found in the publication list below. Dr. Wang provided the original conception of the building extraction project. She proposed the 2D and 3D building extraction from high resolution images for her NSERC Grant and a research grant from Beijing Normal University in 2010. Dr. Wang contributed in the development and formulation of methodology ideas and helped in establishing experimental procedures. She also provided valuable comments, editing and revision on the manuscripts, financial support, and software/hardware/data. Dr. W. Zhan contributed in developing the ideas for Chapter 4. T. Zhao contributed to data preparation for one data source for Chapter 5, and later provided the corresponding descriptions about that data. Other co-authors provided comments on the manuscripts, proofreading or field work assistance.


Acknowledgements

First of all, I thank my supervisor Dr. Jinfei Wang for her invaluable guidance, supervision, encouragement, and support over the past four years. She has kept an open and positive attitude toward my research ideas, and has patiently edited my paper manuscripts. Dr. Wang’s warm-hearted, easy-going generosity as a supervisor and mentor is unforgettable.

I would like to thank my committee members Dr. Micha Pazner and Dr. Phil Stooke. Dr. Pazner has provided me with his unique insights on the connection between science and art. Dr. Stooke has been involved throughout my Ph.D. program, and has provided me with many valuable comments on my proposal and dissertation.

I would also like to acknowledge the contribution of the defense panel. Dr. Marco Van de Wiel always has an open door; his guidance and willingness to listen to my ideas, regardless of their relevance to his own research, is much appreciated. Dr. James Voogt and Dr. Hanping Hong provide me with many usefully comments and suggestions. I am thankful to Dr. Ying Zhang from Natural Resources Canada, who always provided me with positive comments and suggestions to encourage me move forward.

I am grateful to my lab-mates in our Geographic Information Technology & Applications Lab, who have accompanied me in the last four years, sharing research ideas and wonderful lives together. My thanks to Nicholas John Lantz, Francisco Flores De Santiago, Wenfeng Zhan, Autumn Gambles, Ting Zhao, Qin Ma, Peng Wang, Jing Xia, and Wen Dou. Thanks to the students in Beijing Normal University and their supervisor Dr. Chunyan He, for their assistance with my fieldwork in torrid summers.

I also thank the Department of Geography administration team, Dan Shrubsole, Lori Johnson, Angelica Lucaci, Caroline Majeau, Joe Smrekar, Kathy Tang, and Erika Hill, for their support of my research and teaching.

This thesis research was partially funded by an NSERC Discovery Grant awarded to Dr. Wang and a research grant from the State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, China, awarded to Dr. Wang. I would like
to acknowledge the financial support by the University of Western Ontario over the past four years.

Finally and most importantly, I would like to thank my family. My wife, Cheng Qiao, has sacrificed her time to support and encourage me. I would like to express my gratitude and appreciation to my parents for their understanding and support of my life and career choices.
# Table of Contents

Abstract .................................................................................................................................................. ii
Co-Authorship Statement ......................................................................................................................... iv
Acknowledgements ................................................................................................................................. v
Table of Contents ..................................................................................................................................... vii
List of Tables ........................................................................................................................................... xi
List of Figures .......................................................................................................................................... xiii
List of Appendices .................................................................................................................................... xvii
List of Abbreviations, Symbols, Nomenclature ....................................................................................... xviii

Chapter 1: Introduction .............................................................................................................................. 1
  1.1 Building footprint extraction from remotely sensed data. ................................................................. 4
  1.2 Building height estimation from remotely sensed data ................................................................. 8
  1.3 Stereo image matching and digital surface model generation ..................................................... 15
  1.4 Three dimensional building reconstruction from remotely sensed data. ................................. 24
  1.5 Thesis objectives and structure .................................................................................................... 32

References .................................................................................................................................................. 34

Chapter 2: Building Footprint Extraction from High Resolution Remotely Sensed Data and a New Evaluation System .................................................................................................................. 40
  2.1 Introduction ...................................................................................................................................... 40
    2.1.1 A brief review of building extraction methods ................................................................. 41
    2.1.2 A critical review of evaluation methods for extracted buildings ....................................... 43
    2.1.3 A multi-criteria evaluation system ......................................................................................... 50
  2.2 Experiments for building extraction methods ............................................................................... 57
    2.2.1 Study Area and Data Sources ............................................................................................. 57
    2.2.2 The Four Different Building Extraction Methods .......................................................... 60
    2.2.3 Building Extraction Result Maps ........................................................................................ 64
  2.3 Selection and summary of metrics at different levels ...................................................................... 64
    2.3.1 Redundancy Reduction in Shape Similarity Based on Principal Component Analysis... 64
    2.3.2 The Linear Correlation between Positional Accuracy Metrics at Per-building Level... 66
    2.3.3 The Aggregation of Duplicate Metrics Based on the Correlation Analysis. ................... 68
    2.3.4 Summary of Metrics to the Overall Level ....................................................................... 68
  2.4 Results and discussions .................................................................................................................... 71
4.3.4 Building-ground EDM adjustment: buildings on slope ........................................ 122
4.4 Experiments: building-ground EDM implementation ......................................... 127
  4.4.1 Study areas and data sources ......................................................................... 127
  4.4.2 Experiment steps .......................................................................................... 128
4.5. Result: Roof-top point height evaluation and 3D building reconstruction ......... 131
  4.5.1. Performance evaluation: building height error analysis .............................. 131
  4.5.2 Building-ground EDM elevation model in dense urban areas ...................... 134
  4.5.3 The 3D building reconstruction for typical buildings ................................... 136
4.6. Conclusion ........................................................................................................ 138
References .................................................................................................................. 139
Chapter 5: A Multi-criteria Evaluation Method for 3D Building Reconstruction ........ 143
  5.1 Introduction ........................................................................................................ 143
    5.1.1 The Status of 3D Building Reconstruction ................................................ 144
    5.1.2 The Framework to Review 3D Building Evaluation ..................................... 145
  5.2 Methodology ..................................................................................................... 148
  5.3 Experiments and evaluation process .................................................................. 154
    5.3.1 Three dimensional building reconstruction from LiDAR data .................. 154
    5.3.2 Three dimensional building evaluation via the multi-criteria evaluation system ...... 154
  5.4 Results and Discussion .................................................................................... 161
  5.5 Conclusion ........................................................................................................ 166
References .................................................................................................................. 167
Chapter 6: General discussion and conclusions ......................................................... 171
  6.1 Summary .......................................................................................................... 171
  6.2 Conclusions and Contributions ....................................................................... 173
  6.3 Future research ................................................................................................ 176
    6.2.1 Very high-resolution images for 2D building footprint extraction .............. 176
    6.2.2 Multi-view stereo image matching and multi-platform image integration .... 178
    6.2.3 Combining terrestrial sensors with air-borne sensors for 3D building reconstruction ................................................................. 180
    6.2.4 Building information application in natural disaster management .......... 180
References .................................................................................................................. 182
Appendix A: Traditional accuracy assessment for image classification ................. 185
Appendix B: Object-based classification and nearest neighbour classifier ........................................ 188
Appendix C: LiDAR data collection, processing, and related software packages .......................... 197
Appendix D: Principal component analysis (PCA) and factor selection........................................ 208
Appendix E: Ground GPS survey forms and pictures ................................................................. 211
Appendix F: Semi-variogram model ............................................................................................ 215
Appendix G: Spherical harmonic (SPHARM) representation for 3D objects ......................... 220
Appendix H: Matlab codes for 3D objects comparison ............................................................... 226
Appendix I: Copyright releases from publications ...................................................................... 235
References for Appendices ........................................................................................................ 237
CV: Chuiqing Zeng ....................................................................................................................... 238
List of Tables

Table 1-1. Existing commercial 3D building reconstruction systems and research prototypes. Adapted from Wang (2013) ................................................................. 29

Table 2-1. Different metrics derived from the traditional image classification evaluation ................................................................................................................................. 45

Table 2-2. The metrics for object-based shape similarity ................................................................. 47

Table 2-3. The correlations between the metrics at per-scene level ........................................ 70

Table 2-4. The metrics at per-scene level after congregation .................................................. 74

Table 2-5. The summarized index at overall level based on the three components ........... 75

Table 2-6. The corresponding metrics of the four example buildings ..................................... 77

Table 2-7. The metric comparison between public buildings and residential buildings .. 80

Table 3-1. The rooftop elevation errors for different stereo image matching methods .. 107

Table 3-2. Detailed elevation errors for building roof points with elevation larger than 70m ......................................................................................................................... 108

Table 4-1. The minimum elevation in buffers and corresponding difference toward building border ........................................................................................................... 126

Table 4-2. Building height estimated errors from four methods ........................................... 133

Table 5-1. Summary of the three components for accuracy evaluation .............................. 150
Table 5-2. The evaluation result for example buildings with multi-criteria. ................. 162
List of Figures

Figure 1-1. An example of 3D building models in dense urban areas............................. 2

Figure 1-2. The stages of building information extraction from 2D to 3D, from simple to complicated................................................................. 3

Figure 1-3. Building footprint extraction using the snake model based on optical imagery. ........................................................................................................ 6

Figure 1-4. Building footprint extraction from DEM. Adapted from Lafarge et al. (2008). ........................................................................................................... 7

Figure 1-5. Building height estimation based on shadows ........................................... 9

Figure 1-6. Illustration of building height estimation based on single SAR image ....... 12

Figure 1-7. The limitation of current building height estimation from SAR images .... 13

Figure 1-8. Stereo imagery is used to detect above-ground object height. Adapted from Fig. 4. in Soergel et al. (2009) : (a) Optical imagery. (b) SAR......................... 14

Figure 1-9. The general process of stereo image matching and DSM generation ....... 18

Figure 1-10. Epipolar geometry: corresponding set of epipolar lines and their epipolar plane adapted from Szeliski (2010). ............................................................... 19

Figure 1-11. Parallax relationship in a stereo pair and elevation calculation. Adapted from Figure 3.17, Page 156 in Lillesand et al. (2008). ......................................................... 22

Figure 1-12. An example of complicated rooftop with definitions of roof elements ..... 25

Figure 1-13. Examples of some popular roof types .................................................. 26

Figure 1-14. Classification of 3D building modelling methods. Adapted from Wang (2013). ........................................................................................................... 28
Figure 1-15. The procedure of 3D building reconstruction based on an aerial image and LiDAR data. Adapted from Sohn et al. (2012).............................................................................................................. 31

Figure 2-1. The reference and extracted curves and their tangent functions.................. 48

Figure 2-2. Hypothetical examples of different building extraction results compared with the reference........................................................................................................................................ 52

Figure 2-3. A multi-criteria and hierarchical evaluation system for building extraction. 56

Figure 2-4. The study area with reference buildings, ground GPS points and sub-regions. The background is the CIR imagery......................................................................................................... 59

Figure 2-5. The procedure of performance evaluation of the four methods for building footprint extraction.......................................................................................................................... 61

Figure 2-6. A subarea showing the four building extraction results................................. 63

Figure 2-7. The principal component analysis for twelve shape similarity metrics. (a) The biplot of the principal components 1 and 2 with samples.......................................................... 65

Figure 2-8. The scatter plot of the corner point distances and the corresponding building centroid distances for buildings ...................................................................................................... 67

Figure 2-9. Four examples of the extracted buildings by two extraction methods (M3 and M4).................................................................................................................................. 76

Figure 3-1. A conceptual example of rooftop identification............................................. 91

Figure 3-2. The pyramid for the left-right matching with three levels. “Rnd” mean a round of matching. “maxD” is the maximum possible disparity set by users. ............... 94

Figure 3-3. The satellite image of the study area overlaid by the ground surveyed points ........................................................................................................................................... 96
Figure 3-4. The flowchart for DSM generation with the proposed image matching method .............................................................................................................................................................................................................................................................................................. 98

Figure 3-5. The comparison of edge detection methods .............................................................. 100

Figure 3-6. The left-right matching rates for a sample building, including rates for edge matching, intensity matching, and overall matching ................................................................................................................................. 102

Figure 3-7. An example of the stereo image matching process .................................................. 103

Figure 3-8. Visualization of generated DSM and 3D buildings. .............................................. 110

Figure 4-1. The problem of DSM generation from stereo images ........................................... 119

Figure 4-2. The process of building-ground EDM computation with building buffers .................. 124

Figure 4-3. Building-ground EDM model on a slope ................................................................. 125

Figure 4-4. The two study sites and their DSM ........................................................................ 129

Figure 4-5. Ground elevation estimation and building height extraction workflow............... 130

Figure 4-6. An example of building height estimation at Downtown ....................................... 135

Figure 4-7. The 3D reconstruction of three types of buildings ................................................. 137

Figure 5-1. Illustration of 3D building evaluation ................................................................. 152

Figure 5-2. The procedure of 3D building reconstruction from LiDAR data .............................. 156

Figure 5-3. Framework for 3D building evaluation. Note that each component has its evaluation features and comparison functions ...................................................................................................................... 157

Figure 5-4. Random sample points are used to evaluate volume accuracy ............................ 158

Figure 5-5. 3D buildings sphere parameterization and SPHARM ........................................... 159
Figure 5-6. The *point* accuracy evaluation for an example building. ......................... 160

Figure 5-7. Examples of 3D building shape comparison.............................................. 163
List of Appendices

Appendix A: Traditional accuracy assessment for image classification ......................... 185
Appendix B: Object-based classification and nearest neighbour classifier ..................... 188
Appendix C: LiDAR data collection, processing, and related software packages .......... 197
Appendix D: Principal component analysis (PCA) and factor selection ......................... 208
Appendix E: Ground GPS survey forms and pictures .................................................. 211
Appendix F: Semi-variogram model .............................................................................. 215
Appendix G: Spherical harmonic (SPHARM) representation for 3D objects......... 220
Appendix H: Matlab codes for 3D objects comparison .............................................. 226
Appendix I: Copyright releases from publications ......................................................... 235
List of Abbreviations, Symbols, Nomenclature

**DEM**: Digital Elevation Model. A DEM is a digital model or 3D representation of a terrain's surface — commonly for a planet (including Earth), moon, or asteroid — created from terrain elevation data. DEM is a general term for a digital ground elevation layer.

**DSM**: Digital Surface Model. Similar to DEM but more precisely defined, a DSM represents the earth's surface and includes all objects on it.

**DTM**: Digital Terrain Model. In contrast to DSM, a DTM represents the bare ground surface without any objects like plants and buildings.

**LiDAR**: Light Detection And Ranging. LiDAR is a remote sensing technology which measures properties of scattered light to find range and/or other information about a distant target, by illuminating a target with a laser and analyzing the reflected light.

**NDVI**: Normalized Difference Vegetation Index. NDVI is a simple graphical indicator that can be used to analyze remote sensing measurements, typically but not necessarily from a space platform, and assess whether the target being observed contains live green vegetation or not.

**OBIA**: Object-Based Image Analysis. Similarly, “geospatial object based image analysis”, or GEOBIA. Compared with pixel-based image analysis, OBIA implements remote sensing image analysis (i.e. image segmentation, edge-detection, feature extraction, and classification) based on groups of pixels as processing units.

**PCA**: Principal Component Analysis. PCA is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables.

**SAR**: Synthetic Aperture Radar. SAR is a form of radar, characterized its use of relative motion, between an antenna and its target region, to provide distinctive long-term
coherent-signal variations that are used to obtain finer spatial resolution of data, than is possible with conventional beam-scanning means.

**SPHARM**: SPherical HArmonic Representation. It is a unified surface data smoothing, surface parameterization and surface registration technique. Spherical harmonics are a specific set of spherical harmonics that form an orthogonal system, which can be used to decompose 3D objects as a series of coefficients.
Chapter 1: Introduction

Two-dimensional (2D) building footprints and three-dimensional (3D) building models provide important information for natural disaster management, city planning departments, telecommunication companies, and real estate and insurance industries. For example, natural disasters such as earthquakes and floods often have devastating impacts on human lives, economy, and environment (Jayaraman et al., 1997). Acting as the shelter for human beings and as the location of human property, buildings are vulnerable to various disasters; thus, buildings draw concern from academic and public communities. Additionally, reconstructed 3D buildings are of great importance to heritage properties preservation efforts, signal transmission simulation (Stanivuk, 2012) in telecommunication companies, virtual environment in video games and urban war simulators. An example of a reconstructed 3D building model is provided in Figure 1-1. However, it is both expensive and time-consuming to manually collect building footprints and structure data. For large urban areas containing millions of buildings, an automated method is the only reasonable solution. Remote sensing provides a powerful tool that can detect and update building information in an efficient, low-cost, and rapid-response manner.

With remote sensing imagery, building information is generally extracted step by step with increasing level of details (LoD) (Gröger & Plümer, 2012). As illustrated in Figure 1-2, a 2D building boundary is outlined based on the original image; subsequently, detailed 2D roof parts are delineated, building height is extracted, and a 3D building model is extruded. Additional steps can be performed using advanced terrestrial sensors to detect wall facets and render textures for each facet. Furthermore, indoor objects such as furniture and room layout can be reconstructed using hand-hold camera images (Vouzounaras et al., 2011), but interior object reconstruction at LoD4 is beyond the scope of this thesis, as traditional remote sensing techniques cannot obtain information about features inside building structures.
Figure 1-1. An example of 3D building models in dense urban areas
(Photo courtesy of GeoSim: http://www.magicalurbanism.com/archives/48 )
(a) A building in remotely sensed imagery
(b) The extracted building footprint
(c) The building footprint with detailed roof parts
(d) The simple 3D building by extruding the footprint by heights
(e) 3D building with detailed wall facets
(f) 3D building with surface texture

Figure 1-2. The stages of building information extraction from 2D to 3D, from simple to complicated.
1.1 Building footprint extraction from remotely sensed data.

Building boundary (footprint) provides primary information about a building, since a footprint shows the exact position and potential shape of a building. Many studies have been conducted regarding building footprint extraction from remote sensing imagery (Ahmadi et al., 2010; Champion et al., 2010; Lafarge et al., 2008; Michaelsen et al., 2010; Tournaire et al., 2010). Different data sources are used for building footprint extraction. Various data include early aerial images (Huertas & Nevatia, 1988; Irvin & McKeown, 1989), interferometry synthetic aperture radar (InSAR) (Gamba et al., 2000), recent high-resolution satellite imagery (Lee et al., 2003) and related digital surface models (DSM) (Lafarge et al., 2008), light detection and ranging (LiDAR) point clouds (Forlani et al., 2006; Zhang et al., 2006), and terrestrially sensed images (Pu & Vosselman, 2009). To exploit the respective advantages of each data source, data combination is also widely used for building footprint extraction (Simonetto et al., 2005; Sohn & Dowman, 2007; Tupin & Roux, 2003; Turlapaty et al., 2012).

Building footprint extraction methods search for discontinuities in data from a building to the immediate neighbouring ground to find the building’s edges. In an optical image, the colour contrast between a building roof and the ground or surrounding trees is employed to search for building edges. With respect to a DSM and LiDAR, height discontinuity from a roof to the ground is utilized to detect edges. However, in optical imagery, adjacent ground (i.e. paved roads) may have the same colour as the roof, which can lead to spectral confusion between grounds and roofs. A similar problem occurs for height data when a building is surrounded by trees of similar height. As a result, a single data source usually results in low accuracy with regards to extracted building footprints. The combination of spectral (colour) and height data can remove interference so as to maximize extraction accuracies.

Figure 1-3 provides an example of building extraction based on spectral information. The active contour model, also referred to as “snake model” (Kass et al., 1988) because it
moves like a snake during the optimization, is a framework to delineating an object outline from a possibly noisy 2D image. The cost to separate a building from its surroundings, referred to as the “energy” of a snake model, is minimized at corners and edges (Kabolizade et al., 2010). In another example illustrated in Figure 1-4, height information can be used for building footprint extraction. This method compares the DSM discontinuities to the borders by extracting points of interest from each slice and checking the level of coherence between these points and the rectangular shape of the object.

After 2D buildings are extracted, post-processing, including polygon simplification (Zhang et al., 2006) and shape regularization (Sampath & Shan, 2007), are employed in order to generate clean and precise building footprints. In summary, the current literature uses different algorithms based on the various data sources and methods employed. Many building footprint extraction methods have been developed; however, it is difficult to compare the extracted results from different methods comprehensively to report the strength and weakness of methods and provide guidance for method selection. Currently, the most commonly used method is to calculate the intersection between a reference building and a sample building to count the areas of true positive (TP), false positive (FP) and false negative (FN). These areas are used to evaluate building extraction accuracy (McKeown & Cochran, 1999). In another study, current building footprint extraction comparison (Rottensteiner et al., 2013) analyzes the level of accuracy in terms of area completeness, automation level, extraction techniques used, etc. However, this study focuses the comparison of methods from its implementation procedure rather than the extracted building accuracy.
Figure 1-3. Building footprint extraction using the snake model based on optical imagery. Adapted from Kabolizade et al. (2010): (a) building image, (b) an initial contour for snake iteration using the initial seed-selecting criteria, (c) snake result.
Figure 1-4. Building footprint extraction from DEM. Adapted from Lafarge et al. (2008).

(a) A part of a DSM with a proposed rectangle and its slices. (b) Points of interest are detected using a profile simplification algorithm to represent DEM discontinuities on the slices.
1.2 Building height estimation from remotely sensed data

After a 2D building footprint has been mapped, building height is another important parameter to describe buildings. Technically, a building has no constant “height” because a rooftop cannot be completely flat (further described in Figure 1-12). Nevertheless, a building’s average height provides valuable information that roughly describes the volume of a building. Based on one constant height, a building footprint can be extruded and a coarse 3D building model can be built. Therefore, building height is an important component for transition from a 2D building footprint to a 3D building model. In addition, many satellite sensors have limited detection capabilities; often, sensors can only estimate an average height, not including various roof parts heights, for each building.

Building height can be estimated from remotely sensed data via different methods. Building shadows can indicate height (Lillesand et al., 2008), as the ratio of building height to the shadow length is constant in one image depending on the solar zenith angle ($\theta$), as shown in Figure 1-5. If a certain ground object and its shadow length can be measured (i.e. the $h_0$ and $s_0$ in Figure 1-5), then heights of other buildings can be calculated based on their respective shadows. However, shadows vary with the solar position in daytime; more importantly, shadows of buildings can be altered due to the effects of environmental features, such as terrain, adjacent trees, and buildings. In studies of building height estimation with shadows (Irvin & McKeown, 1989), the estimated building height accuracy is low (at 13m level) on high-rise buildings (Shao et al., 2011), which can predict worse results for low-rise buildings due to shadow inaccuracy.
Figure 1-5. Building height estimation based on shadows

\[ \tan \theta = \frac{s_1}{h_1} = \frac{s_0}{h_0} \]  \hspace{1cm} (1-1)
Apart from shadows, single Synthetic Aperture Radar (SAR) imagery also has the ability to detect heights of above-ground objects. Based on a hypothesis of building height in a study area, a corresponding SAR image is simulated. With the assumption of building height changes, a series of SAR images are generated. The best matching between the actual SAR image and one simulated SAR image in the series provides the optimal building height estimation (Brunner et al., 2010; Guida et al., 2010).

Considering the example of a flat-roofed building (Figure 1-6), the layover area of signal superposition from ground, façade, and roof, starts at the image point of the first roof edge and ends at the building wall, whereas on the rear side the ground is occluded by radar shadow of length \( s \). The dimensions of \( d \) and \( s \) on ground are given by (Soergel et al., 2009):

\[
d = h \times \cot \theta_1, \quad s = h \times \tan \theta_s
\]  

(1-2)

Height estimation methods using SAR imagery, however, are still limited by the simple scenario modelling (Guida et al., 2010) and the difficulties surrounding the presence of joined buildings over dense urban areas (Brunner et al., 2010). As illustrated in Figure 1-7, a minimum distance between buildings is required in order to make sure a second building does not interfere with the returned signal. Finally, complicated roof shapes also obstruct building height detection from SAR images.

Furthermore, although LiDAR is often employed to reconstruct 3D buildings with roof structures, the range measurement mechanism of LiDAR (Lillesand et al., 2008) provides a solid theoretical background for building height estimation.

Alternatively, stereo imagery is another method to estimate building heights based on the image correspondences and parallax in high-resolution optical imagery (Lillesand et al., 2008) or SAR images (Soergel et al., 2009). The principle of detecting building height using stereo images is illustrated in Figure 1-8. Stereo images provide imagery from different view angles to assist in object matching and image disparity computation.
The effectiveness of SAR images in dense urban areas is reduced due to intricate triple-bounce and multi-bounce signal scattering among objects. Coarse spatial resolution is an additional concern. Although there are problems with optical stereo imagery for building height estimation (i.e., easily affected by weather condition, occlusion, shadows, and algorithm immaturity), cost-effective stereo optical image matching can generate accurate DSMs over smooth terrain. Stereo optical image matching and its applications for building height estimation will be discussed further in the next section.
Figure 1-6. Illustration of building height estimation based on single SAR image. Adapted from Fig.2 in Soergel et al. (2009), with sensor altitude $H$, height of object $h$, and local viewing angle $\theta$ ($\theta_l$ and $\theta_s$ being the viewing angles at the building locations causing layover and shadow).
Figure 1-7. The limitation of current building height estimation from SAR images. It is adapted from Fig.7 in Brunner et al. (2010): Minimum distance $\Delta_{\text{min}}$ required between two buildings in order that their scattering effects do not interfere.
Figure 1-8. Stereo imagery is used to detect above-ground object height. Adapted from Fig. 4. in Soergel et al. (2009) : (a) Optical imagery. (b) SAR.
1.3 Stereo image matching and digital surface model generation

High-resolution stereo images are one of the major data sources for DSM generation and building height estimation, especially considering the accuracy of derived DSM compared with prohibitively priced LiDAR data. For example, the DSMs generated from Geoeye-1 and Worldview-2 stereo image pairs (Aguilar et al., 2014) give vertical accuracy over the whole study area at about 1 meter. With the advent of new high-resolution satellite sensors such as Pleiades (D'Urso et al., 2010), ZiYuan-3 (Pan et al., 2013), GeoEye-1, and the coming Worldview-3 (DigitalGlobe, 2014), triple-or multi-view stereo images are predicted to become more popular. Stereo image matching over multi-view images is expected to exploit the availability of redundant image information and produce more accurate DSM comparable to that of LiDAR-derived results (Hirschmuller & Bucher, 2010).

Stereo matching is the process of taking two or more images, and estimating a 3D model of the scene by finding matching pixels in the images and converting their 2D positions into 3D depth (Szeliski, 2010). Stereo image matching measures object elevation using the image parallax. The term parallax refers to the apparent change in relative positions of stationary objects caused by a change in viewing position (Lillesand et al., 2008). As the imaging point of view changes, features appear to move to the relative lower elevation features. These relative displacements, referred to as “parallax”, are the basis for three-dimensional viewing of overlapping photographs. The 3D depth map can be further warped on to grids and produce DSMs, using the related spatial reference (Lillesand et al., 2008). The following diagram (Figure 1-9) outlines the steps typically used to generate a DSM from stereo image pairs.

With a given stereo image pair, the process of image orientation is to recover camera positions, restore relative pose and calibrations of the cameras, and build an object-to-image space transformation. Specifically in remote sensing, the rational polynomial
coefficients (RPC) and ground control points (GCP) are popular data sources used for image orientation.

Most high-resolution satellite images distribute RPC files together with stereo image pairs. The RPC algorithm uses the ratio of two polynomial functions to describe the transformation between image coordinates and ground coordinates. The RPC are scene specific coefficients that imitate fraction format of the physical model to describe the imaging geometry and the transformation between the object space and image space (Di et al., 2003). The RPC is essentially a generic form of polynomials. When the denominator is equal to 1, RPC become regular 3D polynomials (Tao & Hu, 2001). The image coordinates and ground coordinates are normalized to the range from -1.0 to 1.0. For one image, the rational functions can be expressed as (Di et al., 2001; Grodecki, 2001):

\[
x = \frac{P_1(X, Y, Z)}{P_2(X, Y, Z)}, \quad y = \frac{P_3(X, Y, Z)}{P_4(X, Y, Z)} \quad (1-3)
\]

\[
P_l(X, Y, Z) = \sum_{i=0}^{m_1} \sum_{j=0}^{m_2} \sum_{k=0}^{m_3} a_{l(i,j,k)} X^i Y^j Z^k, \quad l=1, 2, \ldots, 4, \quad (1-4)
\]

i.e. \( P_1(X, Y, Z) = a_0 + a_1 X + a_2 Y + a_3 Z + a_4 X^2 + a_5 X Y + a_6 X Z + a_7 Y^2 + a_8 Y Z + a_9 Z^2 + a_{10} X^3 + a_{11} X^2 Y + a_{12} X^2 Z + a_{13} X Y^2 + a_{14} X Y Z + a_{15} X Z^2 + a_{16} Y^3 + a_{17} Y^2 Z + a_{18} Y Z^2 + +a_{19} Z^3 \) \quad (1-5)

where \((x, y)\) in Equation (1-3) are normalized image coordinates in pixel (i.e., offset and scaled to [-1, +1]), and \((X, Y, Z)\) are normalized ground coordinates. The polynomials \(P_l,(l=1,2,\ldots,4)\) have the form in Equation (1-3). Usually, the maximum powers of ground coordinates are constrained by \(0 \leq m_1 \leq 3, 0 \leq m_2 \leq 3, 0 \leq m_3 \leq 3,\) and \(m_1+m_2+m_3 \leq 3\). Each \(P(X, Y, Z)\) is a third order, 20-term polynomial. \(1 \leq l \leq 4\), denotes the four coefficients.

The distortions caused by the optical projection can generally be represented by the ratios of first-order terms, while corrections such as Earth curvature, atmospheric refraction,
lens distortion, etc., can be well approximated by the second-order terms. Some other unknown distortions with high-order components, such as camera vibration, can be modeled with the third-order terms (Tao & Hu, 2001). Compared with directly regular polynomial case, using the denominators can always reach a better transformation accuracy (Tao & Hu, 2001), especially when the control points in ground space are not distribute evenly. In most cases, highly accurate GPS receivers are used to collect ground control points (GCP) in study areas. The integration of GCPs into the RPC model can largely improve the stereo model accuracy by providing redundant information and solve the transformation using least square optimization (Tao & Hu, 2002).

To reduce the number of potential correspondences, speed up the matching process, and increase reliability, epipolar images are created based on the image orientation and geometry relationship between camera centres, the ground point, and the image correspondences, as shown in Figure 1-10. In epipolar images, the correspondence of a pixel is only shifted in one image dimension.
Figure 1-9. The general process of stereo image matching and DSM generation
Figure 1-10. Epipolar geometry: corresponding set of epipolar lines and their epipolar plane adapted from Szeliski (2010).

The left and right images are taken at original camera centres $c_0$ and $c_1$ respectively. For a point $p$ on the ground, both images have $p$’s corresponding pixel in the image plane, denoted as $x_0$ and $x_1$. Connecting the two camera centres will intersect with the two image planes at $e_0$ and $e_1$. $x_0$ and $e_0$ forms epipolar line $l_0$ in the left image, whereas $e_1$ and $x_1$ forms the right epipolar line $l_1$. As vectors $\overline{e_1} = (c_1 - c_0)$, $\overline{e_2} = (p - c_0)$, and $\overline{e_3} = (p - c_1)$ are co-planar and they share the same plane with epipolar lines $l_0$ and $l_1$, it can be determined that for the given point $x_0$ in the left image, its corresponding point in the right image should on epipolar line $l_1$ because of the co-planar constraint.
Based on these epipolar images, the most important step in elevation computation is to match the left and right images and find correspondences. Stereo image matching is categorized as intensity-based and feature-based methods (Gruen, 2012). In the **intensity-based** matching, the original or pre-processed image data is used in the form of a matrix of digital number (DN) values. Popular methods include normalized cross-correlation (NCC), least squares matching (LSM), etc. In **feature-based** matching, basic image primitives such as points, edges, and corners are initially extracted from images and then matched (Kim et al., 2001; Suveg & Vosselman, 2004). Compared with image intensity, features are probably more stable with regard to reflectance characteristics, but information is lost during the feature extraction. For dense image matching that must generate elevation values for each pixel, an intensity-based method is appropriate and, as such, is widely used in commercial software.

In an intensity-based matching methodology, the process of image matching is to search for highest intensity similarity between left and right images. For example, the sum of squared differences (SSD) algorithm searches for correspondent pixels based on a moving window. The parallax for a pixel is selected when $E_{SSD}$ is minimized:

$$E_{SSD}(d) = \sum_i [I_l(x_i + d, y_i) - I_r(x_i, y_i)]^2 = \sum_i e_i^2 ,$$  \hspace{1cm} (1-6)

where $I_l$ and $I_r$ are the left and right epipolar images, $i$ is a pixel in the moving window, $(x_i, y_i)$ is the coordinates of pixel $i$, and $d$ is the displacement on $x$ direction (related to parallax). The optimal displacement that minimizes $E_{SSD}$ is the parallax $p$.

From the perspective of computation, stereo algorithms consist of the following four steps: matching cost computation, cost (support) aggregation, parallax computation and optimization, and parallax refinement. Normalized cross-correlation (NCC) (Gruen, 1985) and sum-of-squared-differences (SSD) compute pixel similarity based on the image intensity. In contrast, some similarity measures are insensitive to image gain and bias such as hierarchical mutual information and census transform (Hirschmuller, 2008),
which converts each pixel inside a moving window into a bit vector representing which neighbours are above or below the central pixel. Additionally, local (window-based) algorithms, where the parallax computation at a given point depends only on intensity values within a finite window, usually make implicit smoothness assumptions by aggregating support. Global algorithms, on the other hand, make explicit smoothness assumptions and then solve a global optimization problem.

After the parallax $p$ is determined for a pixel, a simple inverse relationship between flight height $H$ and parallax $p$ can be built, as illustrated in Figure 1-11.

Given a pair of stereo images with a matched corresponding point in left and right (epipolar) images with $x$ coordinates $x_a$ and $x'_a$ respectively, the height of this point ($h$) can be calculated as follows (Lillesand et al., 2008):

$$ h = H - \frac{Bf}{p_a}, \quad p_a = x_a - x'_a $$

(1-7)

where $H$ is the given flying height, $f$ is focal length, $B$ is the baseline width, and $p_a$ is the image parallax (Blaschke, 2010).

However, the matching of stereo image pairs has been found to perform poorly in urban areas with discontinuous features such as tall buildings. In this thesis, tall buildings are equivalent to “high-rise” buildings in building engineering. Generally speaking, buildings with eight or more storeys are “high-rise” buildings, buildings with 4-7 storeys are “mid-rise” building, whereas buildings with 1-3 storeys is “low-rise” building, or “low buildings”. Recent studies (Aguilar et al., 2014) report a larger elevation error in urban areas than that encountered in bare ground. The actual quality difference between LiDAR data and stereo image derived DSM is concealed by accuracy metrics. To improve the DSM derived from stereo images, extra information/constrains are required in order to reduce matching candidate and improve the matching accuracy.
Figure 1-11. Parallax relationship in a stereo pair and elevation calculation. Adapted from Figure 3.17, Page 156 in Lillesand et al. (2008).

The $L$ and $L'$ are the locations where photos are taken. $f$ is the focal length. $o$ and $o'$ are the photo centers. $a$ and $a'$ are the locations of a correspond point ($A$ on the ground) in left and right photo. $H$ is the flight height. $B$ is the baseline between the two image centres.
Gruen (2012) listed many constrains including: (a) epipolar geometry; (b) multi-view matching; (c) limits on the magnitude of changes in parallax; (d) a priori modelling of objects (coarse description of object); (e) hierarchical “coarse-to-fine” strategies; (f) “best-first” strategies, using features sequentially according to the relevance of their information content; (g) “thin-to-thick” or “thick-to-thin” strategies; and (h) observation of behaviour of parallaxes. In (Wu et al., 2012), triangle constraint is used, which suggests that points inside a pair of corresponding triangles will find correspondences inside triangles. The image matching is then conducted hierarchically from “coarse-to-fine” levels.

A building’s footprint is a widely used data source and is commonly employed in many applications. A building footprint provides valuable prior-knowledge about a building’s position. For studies focused on assessing accuracy of building elevations, a building footprint can serve as constraint in order to narrow down the matching areas and reduce matching candidates. It is also useful to further estimate building height from DSM after matching; however, limited current research exists concerning how to exploit footprint information during the matching process.

Other challenges in stereo image matching concerns the occurrence of occlusion and shadowed areas. Most image matching algorithms/software use interpolation to fill the invalid matching areas. Specifically for buildings, the interpolated DSMs fill in failed areas around buildings, as immediate neighbouring areas are always affected by occlusion or shadow, resulting in successful matching. It is difficult to estimate accurate building height using such interpolated DSMs. A straightforward solution is counting different zonal statistic variants for given building footprints (Tack et al., 2012), such as average or median height, where the statistic with the best approximation of the correct building height is selected. However, the selection of the best variant for height requires prior knowledge about each building’s height, which contradicts the goal (height
estimation). Better solutions are required to address issues related to interpolated inaccurate DSMs and estimation of building height, where it is preferable to directly estimate building height using the DSM.

1.4 Three dimensional building reconstruction from remotely sensed data.
As introduced in the previous section, a building does not have a constant height. Generally, building roofs can have many parts such as ridges, hips, and eaves, as shown in Figure 1-12. One building height measurement extrapolated to include an entire roof is not enough to support detailed 3D reconstruction. To accurately describe a 3D building model, the heights of each roof part need to be estimated individually; consequently, advanced sensors and very high resolution data are required for detailed 3D modelling. For example, LiDAR data with high point density and aerial photos with centimeter spatial resolution are popular for building reconstruction applications.

Typical roof types have been investigated in the current literature, such as the flat, gable, hip, and shed roofs identified in Figure 1-13. A successful reconstruction of typical roofs provides a solid basis for complicated roof reconstruction, as complicated roofs are based on typical roofs and often contain individual components derived from typical roof types. Pre-defined roof types can simplify the reconstruction task and provide a prior knowledge for the automatic reconstruction. A roof type assumption can largely improve the reconstruction accuracy, and is thus widely used in studies (Henn et al., 2013; Huang et al., 2013).
Figure 1-12. An example of complicated rooftop with definitions of roof elements (Photo courtesy of Wikipedia: http://en.wikipedia.org/wiki/File:Roof_diagram.jpg)
### Roof Types

<table>
<thead>
<tr>
<th>Roof types</th>
<th>Conceptual roof shapes</th>
<th>Example roofs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat</td>
<td><img src="image1" alt="Flat Roof Diagram" /></td>
<td><img src="image2" alt="Flat Roof Example" /></td>
</tr>
<tr>
<td>Gable</td>
<td><img src="image3" alt="Gable Roof Diagram" /></td>
<td><img src="image4" alt="Gable Roof Example" /></td>
</tr>
<tr>
<td>Hip</td>
<td><img src="image5" alt="Hip Roof Diagram" /></td>
<td><img src="image6" alt="Hip Roof Example" /></td>
</tr>
<tr>
<td>Shed</td>
<td><img src="image7" alt="Shed Roof Diagram" /></td>
<td><img src="image8" alt="Shed Roof Example" /></td>
</tr>
</tbody>
</table>

Figure 1-13. Examples of some popular roof types.
Automatic reconstruction of digital 3D building models based on 2D building footprints remains a challenging task. Some 3D building methods simply assign a constant height to a 2D building footprint and extrude the building up (Lafarge et al., 2008); essentially, the resultant extruded building is not a 3D, but a 2.5D model. Most methods are developed on LiDAR data for the building roof reconstruction (Haala & Kada, 2010; Khattak et al., 2013; Kong et al., 2013). A systematic review of current 3D building reconstruction is studied in Wang (2013), where reconstruction methods are roughly categorized based on their data as image-based, LiDAR-based, or image-LiDAR fusion methods. This classification framework is introduced in Figure 1-14 and current popular software used for 3D building reconstruction is provided in Table 1-1.

Furthermore, algorithms required to automatically reconstruct 3D buildings can be divided into model-driven and data-driven methods. In model-driven methods, typical roof types are predefined with input point clouds or DSMs fitted to the predefined roof types for modelling (Henn et al., 2013; Huang et al., 2013). In data-driven methods, point clouds and digital surface models (DSMs) are segmented and grouped, features are recognized, and 3D models are built accordingly (Lafarge & Mallet, 2012; Zhang et al., 2012). As in the example provided in Figure 1-15, aerial image and LiDAR data are used to cluster heights, compose planes, detect edges, reconstruct facets, and conduct post-processing.
Figure 1-14. Classification of 3D building modelling methods. Adapted from Wang (2013).
Table 1-1. Existing commercial 3D building reconstruction systems and research prototypes. Adapted from Wang (2013)

<table>
<thead>
<tr>
<th>System</th>
<th>Developer/researcher</th>
<th>Input data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC-Modeler</td>
<td>CyberCity AG &amp; ETH, Zurich</td>
<td>Calibrated stereo pan of aerial images</td>
<td>Semi-automated photogrammetric 3D reconstruction system</td>
</tr>
<tr>
<td>inject</td>
<td>Inpho GmbH &amp; Bonn University, Germany</td>
<td>Calibrated single, stereo, or multiple overlapping aerial images</td>
<td>Semi-automated Constructive Solid Geometry based approach</td>
</tr>
<tr>
<td>Ascender</td>
<td>University of Massachusetts</td>
<td>Calibrated multiple aerial (nadir and oblique) images</td>
<td>Automated 3D building model reconstruction</td>
</tr>
<tr>
<td>SiteCity</td>
<td>Digital Mapping Laboratory, CMU,</td>
<td>Calibrated multiple aerial (nadir and oblique) images</td>
<td>Semi-automated photogrammetric 3D reconstruction system</td>
</tr>
<tr>
<td>ImageModeler</td>
<td>RealViz &amp; INRIA, France</td>
<td>At least two photos taken from different positions</td>
<td>Accurate 3D measurement and modelling from photos</td>
</tr>
<tr>
<td>PhotoBuilder</td>
<td>Oxford University, UK</td>
<td>Uncalibrated two or more photos</td>
<td>Vanishing points based method to 3D reconstruction</td>
</tr>
<tr>
<td>Nverse Photo</td>
<td>Precision Lightworks, USA</td>
<td>Two or more aerial images</td>
<td>A series of plug-in components</td>
</tr>
<tr>
<td>Shape Capture</td>
<td>ShapeQuest Inc. &amp; NRC, Canada,</td>
<td>Single or more photos</td>
<td>Accurate 3D measurement and modelling from single or more photos</td>
</tr>
<tr>
<td>PhotoModeler</td>
<td>Eos Systems, Canada</td>
<td>Single or more photos</td>
<td>Accurate 3D measurement and modelling from single or more photos</td>
</tr>
<tr>
<td>PhotoGenesis</td>
<td>Plenoptics Ltd, UK</td>
<td>Uncalibrated single or more photos</td>
<td>Semi-automated model-based 3D reconstruction system</td>
</tr>
<tr>
<td>Photosynth</td>
<td>Microsoft</td>
<td>Internet photos</td>
<td>Sparse 3D model generation for navigating images in 3D space</td>
</tr>
<tr>
<td>Pix4UAV</td>
<td>Pix4D, Switzerland</td>
<td>Aerial images</td>
<td>Automatic 3D model generation from aerial images</td>
</tr>
</tbody>
</table>
Table 1-1 (Continue). Existing commercial 3D building reconstruction systems and research prototypes.

<table>
<thead>
<tr>
<th>System</th>
<th>Developer/researcher</th>
<th>Input data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>Apple, USA</td>
<td>Aerial images</td>
<td>Automatic 3D model generation from aerial images</td>
</tr>
<tr>
<td>Edgewise</td>
<td>ClearEdge3D, USA</td>
<td>Range data</td>
<td>3D modelling using range data</td>
</tr>
</tbody>
</table>

One reoccurring issue in current 3D building reconstruction is that often, rooftops alone are reconstructed while walls are assumed to be featureless (Haala & Kada, 2010). The evaluation process for reconstructed 3D buildings has not been thoroughly discussed, with most studies interested in developing sophisticated reconstruction methods. Currently, evaluation methods for 3D building reconstructions are derived from 2D building footprint evaluation methods. For example, 3D buildings are evaluated based on building roof completeness, topological consistency between reference and sample 3D buildings, and geometric accuracy in XY and Z directions (Rottensteiner et al., 2013). Often, building wall correctness and shape similarity are ignored.
Figure 1-15. The procedure of 3D building reconstruction based on an aerial image and LiDAR data. Adapted from Sohn et al. (2012): (a) aerial image, (b) LiDAR data, (c) height clustering, (d) plane clustering, (e) intersection line extraction, (f) step line
1.5 Thesis objectives and structure

The objectives of this thesis are to:

1. Implement current popular building extraction methods, and develop a systematic framework to evaluate extracted building footprints thoroughly, with the focus on building shape similarity and metric redundancy removal.
2. Develop a building height estimation method from a Digital Surface Model (DSM) derived from stereo imagery to search for building height in complicated and inaccurate DSMs.
3. Develop an advanced building height extraction method from stereo imagery constrained by building footprints, in order to overcome the defects of current matching methods that perform poorly on dense urban areas and tall buildings.
4. Reconstruct 3D buildings and develop a new multi-criteria system to evaluate the reconstruction accuracy from different perspectives in a true 3D environment.

Two study sites are employed to develop methods for different building information extraction and evaluation from remotely sensed data. The two study areas are: (a) the campus of the University of Western Ontario, London, Ontario, Canada, and (b) the campus of Beijing Normal University, Beijing, China. The reference data are ground GPS-surveyed absolute elevation and laser rangefinder-measured building relative heights.

Specifically, to investigate a building in remote sensing imagery, the first step is to identify its location. In Chapter 2, the main objective is to implement different extraction methods for building footprints and to evaluate the results in terms of accuracy. Four popular building footprint extraction methods are implemented for comparison between
different data sources. The resultant identified building location is a prerequisite for building height estimation; In Chapters 3 and 4, methods were investigated for building height extraction. In Chapter 3, a method is proposed to use popular optical satellite stereo imagery to extract building roof elevation directly. In Chapter 4, it was discussed about building ground elevation estimation based on a DSM derived from widely used commercial software and building footprints derived from the methods presented in Chapter 2; then building height can be calculated from roof and ground elevation contrast. In Chapter 5, 3D building reconstruction is discussed, and effective accuracy evaluation methods for reconstructed 3D buildings derived from a range of methods and data are explored.
References


Chapter 2: Building Footprint Extraction from High Resolution Remotely Sensed Data and a New Evaluation System*

2.1 Introduction

Building footprint detection from remotely sensed data is of great importance to disaster (earthquake, flood or fire) management, real estate industry, homeland security and many other applications (Awrangjeb et al., 2010). Building footprint extraction is a well-developed topic that has been explored in many studies (Awrangjeb et al., 2010; Blaschke, 2010; Haala & Brenner, 1999). Current automatic methods for building extraction reaches accuracy at 90% level (Rottensteiner et al., 2013); however, it cannot reach a perfect 100% accuracy. The main reasons behind the deficiency are scene complexity, incomplete cue extraction and sensor dependency (Sohn & Dowman, 2007). Considering the large number of extraction methods but lack of standard techniques to evaluate them, the evaluation methods become very important. Although there are several studies (Awrangjeb et al., 2010; Lee et al., 2003; Möller et al., 2007; Rutzinger et al., 2009; Shan & Lee, 2005; Zhan et al., 2005) about the evaluation of building extraction, most of them still concentrate on the image classification accuracy, rather than a framework to thoroughly evaluate the extraction results. In one study an evaluation system is defined (Song & Haithcoat, 2005); however, there is no experiment to empirically verify the system. Another problem with the evaluation criteria is the redundancy of metrics. Most studies list related metrics (e.g. 15 metrics in Awrangjeb et al. (2010)), but cannot provide a clear picture about the performance of a given extraction method.

This study provides a brief review of building footprint extraction methods, and a detailed review of current evaluation metrics for building extraction. The objective of this study is to organize current building evaluation metrics, to reduce the redundancy, and to develop an evaluation system for building extraction accuracy assessment.

2.1.1 A brief review of building extraction methods

Aerial photographs and high-resolution optical satellite images are typical data sources for the extraction of building’s boundaries. Various methodologies have been developed based on the concepts and characteristics of buildings, including edge detection, image classification, Hough transformation, seed growing, geometric, photometric and structural analyzes, and spectral, structural and contextual analyzes (Sahar et al., 2010). In order to tackle the complicated context of buildings in urban areas, other methodologies such as snake model (Ahmadi et al., 2010; Kass et al., 1988) and energy function (Tournaire et al., 2010), have been introduced to building footprint extraction. Recently, buildings have been extracted directly from 3D virtual environment by a template matching technique for height estimation and potential building segment detection (Turlapaty et al., 2012). Optical images alone, however, can only obtain limited spectral information of ground features in visible bands. It is difficult to differentiate ground features with similar spectral characteristics (e.g., buildings and other man-made features). Spatial information from the high-resolution images increases the separability between buildings and non-buildings but still cannot resolve incidences of spectral confusion. As a result, auxiliary information, such as vector parcel geometries and their attributes from Geographic Information Systems (GIS) (Sahar et al., 2010) or high-resolution interferometric SAR (InSAR) data (Wegner et al., 2011), can benefit and improve the building extraction.

In recent decades, the advent of light detection and ranging (LiDAR) data has challenged the conventional building footprint extraction methods. The ability to detect distances (Lillesand et al., 2008), as well as the high spatial resolution, gives airborne LiDAR an advantage in extracting building footprints accurately. LiDAR can be used independently to extract building footprints via height, size, and shape information (A Visual Learning
By tracing and regularization of building boundaries from raw LiDAR point clouds, some studies separate building and non-building LiDAR points via slope (Sampath & Shan, 2007) or morphological filtering (Zhang et al., 2006), segmenting LiDAR points that belong to the same building, tracing building boundary points, and regularizing the boundaries. These studies report high accuracy on building footprint extraction. Furthermore, LiDAR is even applied to roof plane segmentation and roof model reconstruction via level set approaches (Kim & Shan, 2011).

Finally, optical imagery and LiDAR data are integrated for more accurate building footprint extraction. LiDAR data contains height and intensity but does not provide colour information in contrast; optical imagery has colour bands but lacks height information. As a result, these two complementary data are combined for better recognition of buildings. In some studies, LiDAR is used to extract coarse building boundaries and then overlaid with optical image to retrieve precise building boundaries (Dong et al., 2008; Ekhtari et al., 2009). Alternatively, a group of LiDAR points are clustered as an isolated building object based on point height similarity and homogenous normalized difference vegetation index (NDVI) derived from optical imagery (Awrangjeb et al., 2010; Sohn & Dowman, 2007). A recent effort aims to separate buildings from trees in complex scenes with hilly terrain and dense vegetation (Awrangjeb et al., 2012).

Another trend in building footprint extraction is to utilize optical imagery with higher spatial resolution. This trend is accompanied by the increasing impact of relief displacement on aerial images (Avery & Berlin, 1992) and its considerable effect on building boundary extraction. Currently, little attention has been paid to correct this displacement (Chen et al., 2007; Zahran, 2009; Zhou et al., 2004); few studies have taken true ortho-rectification into consideration before extracting building footprints. A previous study had developed a method to correct the relief displacement of tall objects from ortho-rectified aerial images (Lehrbass & Wang, 2012).
The current building footprint extraction methods use data from mixtures of different sources, various algorithms, and feature different degrees of human interference. Most studies propose methods and evaluate results using inconsistent criteria. This inconsistency between evaluation methods impedes the comparison of building footprint extraction. Therefore, a consistent evaluation system is required.

### 2.1.2 A critical review of evaluation methods for extracted buildings

**A. Matched Rates**

Building footprint extraction results have been evaluated by different strategies in many studies. The most popular evaluation method is transplanted from traditional image classification accuracy assessment (Foody, 2002). In an evaluation, two classes are considered: buildings (or the object) and the background. The extracted objects are evaluated against the reference objects. As a result, four different parameters (McKeown & Cochran, 1999), true positive \(TP\), the common area of extracted objects and reference objects), false positive \(FP\), the area of extracted objects but not reference objects), false negative \(FN\), the area that belongs to the reference but not the extracted result), and true negative \(TN\), the area that belongs to neither the extracted result nor the reference), are derived from the reference image and the building extraction image. Different metrics are developed for evaluation as defined in Table 2-1.

The Completeness, also known as producer’s accuracy (Foody, 2002) or detection rate (Song & Haithcoat, 2005), represents the percent of the buildings area being correctly detected with respect to the reference data; while the Correctness, referred to as user’s accuracy (Foody, 2002) and overlap (Shan & Lee, 2005), describes the percentage of correctly detected area over the total area of extracted buildings. On the other hand, Omission error (Song & Haithcoat, 2005) is the percent that is not detected from the reference, while Commission error (Song & Haithcoat, 2005) is the incorrectly detected part within the extracted buildings. Quality (Lee et al., 2003; McKeown & Cochran, 1999; Rutzinger et al., 2009), similar to fitness (Shan & Lee, 2005), is a metric that
combines the Completeness and Correctness. Kappa coefficient (Zhan et al., 2005) takes into account the agreements contributed by chance. Branching factor and Miss factor (Lee et al., 2003; McKeown & Cochran, 1999) describe two types of possible mistakes in the extraction: the former indicates the rate of incorrectly labeled building pixels, the latter gives the rate of missed building pixels. Underlap and extralap (Shan & Lee, 2005) correspond to Branching factor and Miss factor, but they are slightly different on the definition. The metrics listed are originally used in the image classification at pixel level, they are later adapted in feature-based applications (Awrangjeb et al., 2010; Zhan et al., 2005). Metrics in the second row in Table 2-1 with higher values indicate better extraction performance; metrics in the fourth row with lower values correspond to better performance. Abbreviations are used: Comp denotes Completeness, Corr denotes Correctness, Omiss\(_e\) is Omission error, Comm\(_e\) is Commission error, \(Q\) represents Quality, Mf is Miss factor, Bf is Branching factor, and Kappa represents Kappa coefficient. More details about traditional image classification accuracy assessment are described in Appendix A.

**B. Shape Similarity**

Shape similarity is another category of metrics which describes to what extent an extracted object is similar to the reference with respect to their shapes. The shape similarity is a subjective and not well-defined problem. For a given reference and an extracted building, matching status cannot be well-defined except for the identical case (the extracted shape is exactly the same as the reference). In current literature, two types of strategies have been developed to measure objects’ similarity. One is to evaluate similarity of the boundary (or “contour”) between an object and a reference; another strategy compares an object and its neighbouring features at image level with image related similarity.
Table 2-1. Different metrics derived from the traditional image classification evaluation

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Completeness</th>
<th>Correctness</th>
<th>Quality</th>
<th>Kappa*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>( \frac{TP}{TP+FN} )</td>
<td>( \frac{TP}{TP+FP} )</td>
<td>( \frac{TP}{TP+FN+FP} )</td>
<td>( \frac{n \times (TP+TN) - pr}{n^2 - pr} )</td>
</tr>
<tr>
<td>Metrics</td>
<td>Omission error</td>
<td>Commission error</td>
<td>Branching factor</td>
<td>Miss factor</td>
</tr>
<tr>
<td>Definition</td>
<td>( 1 - \frac{TP}{TP+FN} )</td>
<td>( 1 - \frac{TP}{TP+FP} )</td>
<td>( \frac{FP}{TP} )</td>
<td>( \frac{FN}{TP} )</td>
</tr>
</tbody>
</table>

*in Kappa, \( n=TP+FP+TN+FN \), and \( pr = (TP + FP)(TP + FN) + (TN + FP)(TN + FN) \)
On the boundary level, topological relations such as overlapping, disjoint, and covering for each pair of objects in reference and extracted maps are used to define a metric of similarity (Dungan, 2006). *Corner positional difference, Perimeter difference* and *Area difference* are defined in (Song & Haithcoat, 2005). These differences describe the absolute (corner, perimeter, and area) differences from extracted buildings to the reference, divided by the reference. In contrast to difference, the ratios of the above measures between the extracted building to the reference building (Möller et al., 2007) are another way to describe similarity. To ensure that the ratio is between 0 and 1, the ratio of the minimum to the maximum of a given shape descriptor (i.e. ratio of Area is referred to as r(Area)), is designed (Zhan et al., 2005). Different similarity measures at the building boundary level are summarized in Table 2-2.

Another metric “Tangent Space Representation (TSR)” (Latecki & Lakämper, 2000) is developed to evaluate the object boundary by comparing their outline difference step by step. For each line segment in a shape, the derivative of the segment, called “tangent function”, is calculated and connected consecutively to form a line. As shown in Figure 2-1, two shapes are displayed on the left: the reference in Figure 2-1(a) and the extracted shape in Figure 2-1(b), with the generated tangent function on the right, respectively. The lengths of the two shapes are rescaled to 1 respectively and the tangent function difference between these two lines is used to measure the similarity.

On the image-level, *Moment-Derived Shape Similarity* (Song & Haithcoat, 2005) is designed to measure the similarity. That is because geometric moments can provide an equivalent representation of an image in the sense that an image can be reconstructed from its moment. The moment-based similarity is the Euclidean distance in a space defined by the first two moments (Hu, 1962) as the two axes. In addition, in (Skerl et al., 2006), mutual information (*MI*) and its derivatives, *Joint entropy, Entropy correlation coefficient*, and *Correlation ratio* are used as image similarity metrics, with most of them reported to have similar performances.
Table 2-2. The metrics for object-based *shape similarity*

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Area Difference</th>
<th>Perimeter Difference</th>
<th>Mean ((\bar{r}(Area)))</th>
<th>Std ((\bar{r}(Area)))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>[ \frac{abs(Ae - Ar)}{Ar} ] (i)</td>
<td>[ \frac{abs(Pe - Pr)}{Pr} ] (i)</td>
<td>[ \frac{\min(Ae, Ar)}{\max(Ae, Ar)} ] (i)</td>
<td>[ \frac{\min(Ae, Ar)}{\max(Ae, Ar)} ] (i)</td>
</tr>
</tbody>
</table>

Note: \(Ae\) and \(Ar\) are the areas of the extracted and reference buildings, respectively; while \(Pe\) is the perimeter of the extraction building and \(Pr\) is that of the reference. \(min\) is the minimum and \(max\) is the maximum. \(Mean\) is the average value for the input samples, \(Std\) is the standard deviation. \(i\) represents the \(i\)th building sample for the buildings under evaluation.
Figure 2-1. The reference and extracted curves and their tangent functions. (a) The reference curve, with the curve on the left and its tangent function on the right. (b) The extracted curve and its tangent function.
C. Positional Accuracy

Positional accuracy, also referred to as geometric position accuracy (Awrangjeb et al., 2010) or location accuracy, is another group of metrics that evaluate the geometric position difference between corresponding points of the extracted objects and the reference. Because of the limitations on current data sources and methodology, together with perceived unimportance to applications, positional accuracy is rarely mentioned in current evaluation systems. Root mean squared error (RMSE) is used to measure the positional errors between points in the reference and extracted maps (Song & Haithcoat, 2005). Usually the corner points of buildings are used as the points for evaluation. Furthermore, in Zhan et al. (2005), the Euclidean distance between the centers of the mass of an extracted object and the corresponding object is used as another metric to measure positional accuracy. Both mean and standard deviation are useful for evaluation.

In summary, evaluation strategies for building extraction methods from the current literature are summarized as three different categories based on their evaluation purpose: matched rate, shape similarity, and positional accuracy. There are other ways to group those evaluation methods too. Pixel-based and object-based evaluation methods are commonly adapted in many studies (Awrangjeb et al., 2010; Rottensteiner et al., 2005; Rutzinger et al., 2009). Evaluation methods labeled as pixel level or object (or “building”) level use different spatial data models: a raster image or vector polygons; however most metrics can be calculated in either data model. Moreover, evaluation methods can also be grouped as local and global (Möller et al., 2007) from the perspective of scales. The local level refers to the single object level, while the global level analyzes all objects in the entire image. Finally, the evaluation system can be built at different levels. Three levels (number-based, area-based, and shape similarity indices (Song & Haithcoat, 2005)) of the indices are used depending on how rigorous the requirements of accuracy assessment are for the desired application.

The problem with the current evaluation metrics can be summarized as follows: (1) most evaluation methods concentrate on the traditional classification accuracy, described by metrics such as Correctness, Completeness, and Quality; the other two components of the
evaluation, the shape similarity and positional accuracy, have not attracted enough attention. (2) The shape similarity component, primarily described by area and perimeter ratio is not well-developed; it requires further systematic and thorough analysis in order to be used to describe the similarity of complex shapes. (3) Various metrics are reported in many studies; however, they are not well-organized. It is difficult for users to assess a method with more than ten metrics at the same time when the judgment is based on many separate variables. (4) The correlation between different metrics has not been investigated. Strong correlation between some of the reported metrics indicates redundant information and conceals the nature of evaluation methods.

2.1.3 A multi-criteria evaluation system

A. Assumptions

The evaluation process can be divided into two steps. The first step is the registration of buildings between reference data and the extracted results; the second step is to compute metrics and conduct evaluation. There are efforts involved in automatic implementation and multi-detection consideration (Awrangjeb et al., 2010) in the registration step. The importance of such efforts because they facilitate the evaluation process is acknowledged; however, the focus of this study is on the second step regarding how to develop an evaluation system. To clarify the expression for connected buildings and avoid topological mismatch between reference and the extracted buildings, a semi-automatic registration method is used in this study. The building registration is automatic for “one-to-one” (including zero-to-one and one-to-zero) relations between the extracted buildings and reference buildings. “Many-to-many” relations occur infrequently and only when there are topological mismatch between reference buildings and the extracted buildings (19 out of 761 buildings in this study). “Many-to-many” relations are clarified by a manual method which applies the topological clarification (Rutzinger et al., 2009) via splitting and merging extracted buildings.

The definition of $TP$ is important and there are different ways to calculate $TP$. One popular approach is setting a threshold for the percent of overlap between the extracted and reference buildings (Rottensteiner et al., 2005; Zhan et al., 2005). In another study
(Song & Haithcoat, 2005) no threshold is set, which means as soon as an overlap between the two buildings is detected, that building will be used for validation. Considering that the threshold is subjective, in (Rutzinger et al., 2009) a Point-in-Polygon test is used to determine $TP$. In this study, the $TP$ definition is adapted from Song and Haithcoat (2005) and no percent threshold is set if overlap exists. A minimum size of a building ($4m^2$) is set to avoid obvious mistakes.

**B. The Demand for a Multi-criteria Evaluation**

Although the popular *matched rate* method is effective to evaluate the building’s matching accuracy at pixel level, *matched rate* alone is not sufficient for a complete evaluation for buildings. As shown in Table 2-2, there are three different extracted buildings compared with the reference. In Table 2-2 (b), the courtyard is displaced to two possible locations: A or B; case A and B lead to the same *matched rate* result, but they are two different shapes. In (c), a building is rotated because of image registration problem although it has the same shape as that in the reference; the *matched rate* is quite low for this matching. In (d), a building is extracted with similar size and location but with irregular boundaries; although this extraction will give a high *matched rate*, the perimeter is larger because its jagged boundaries and the shape will be quite different if it is used in 3D reconstruction. In this scenario, the corner positions are changed.
Figure 2-2. Hypothetical examples of different building extraction results compared with the reference. (a) Reference building boundary (a building with courtyard); (b) extracted result 1: an extracted building with the courtyard displaced to A or B; (c) extracted result 2: a rotated building; (d) extracted result 3: a polygon with irregular borders.
The demand of a multi-criteria evaluation system is twofold. On one hand, the commonly used user’s and producer’s accuracies are not able to completely evaluate the performance of building extraction, because other aspects (e.g., the shape similarity and positional accuracy) are not effectively described. Performance on the other aspects is increasingly important in applications that demand more building details such as detailed 3D reconstruction. On the other hand, the accuracy of extracted buildings has increased remarkably with advanced methods and higher-resolution imagery in recent years. For example, recent extraction methods can achieve the traditional user’s and producer’s accuracy at 90% level (Aldred & Wang, 2011; Huang & Zhang, 2012), which leaves limited room for further comparison of those methods. Therefore, a multi-criteria evaluation is expected to meet both demands.

C. The Selection of Metrics for Evaluation

Among all the proposed matched rate metrics, connections can be derived from their definitions. From Table 2-1, the connection between different metrics built from TP, FP, and FN are summarized as follows:

\[
\frac{\text{Comp} + \text{Omiss}_e}{\text{Comp} \cdot \text{Corr}} = 1; \quad \frac{\text{Corr} + \text{Comm}_e}{\text{Comp} \cdot \text{Corr}} = 1
\] (2-1)

\[
Q = \frac{\text{Comp} \cdot \text{Corr}}{\text{Comp} + \text{Corr} - \text{Comp} \cdot \text{Corr}}
\]

\[
Q = \frac{1}{\text{Bf} + \text{Mf} + 1}
\] (2-2)

\[
\frac{1}{\text{Omiss}_e} + 1 = \frac{1}{\text{Mf}}; \quad \frac{1}{\text{Comm}_e} + 1 = \frac{1}{\text{Bf}}
\] (2-3)

From the listed relations, \(\text{Comp}\) and \(\text{Omiss}_e\) describe the two sides of a coin; so do the \(\text{Corr}\) and \(\text{Comm}_e\). \(\text{Omiss}_e\) relates to \(\text{Mf}\); while \(\text{Comm}_e\) and \(\text{Bf}\) share the same relation. Furthermore, \(Q\) is a metric by integrating both \(\text{Comp}\) and \(\text{Corr}\) in its definition. Considering the relations between \(Q\) and other metrics directly or indirectly, \(Q\) represents most of the above metrics. In this study, \(\text{Mf}, \text{Bf}, \text{Comm}_e,\) and \(\text{Omiss}_e\) are discarded to avoid redundant information, while \(\text{Comp}, \text{Corr}\) and \(Q\) are computed for further analysis. \(\text{Kappa}\) is a metric which is suitable to evaluate all buildings at image level.
Apart from the shape similarity metrics mentioned in the previous Section 2.1.2, many shape indices are designed in eCognition 8.7 to measure an object’s shape (Trimble, 2011). These indices include Area, Perimeter (Peri), Border index (bdr_idx), Asymmetry (Asymm), Density (Dst), Compactness (Cmpt), Length/width (L/W), Main direction (Dir), Elliptic fit (Elp_fit), Rectangle fit (Rect_fit), shape index (Shp_idx) and Roundness (Rnd). Area and Perimeter are the primary metrics to measure a shape. The Asymmetry describes the relative length of an image object, in comparison with a regular polygon. Compactness is calculated by the product of the length and the width, and divided by the number of pixels in an object. Length/width describes whether the object is close to a square (when it equals to 1). Main direction represents the major direction of the object. Elliptic fit and Rectangular fit calculates to what extent the object can be fitted in an ellipse or a rectangle. The Roundness describes how similar an object is to an ellipse.

These shape indices can be easily converted to shape similarity metrics by computing the ratio between an extracted building and a reference building, according to the definition in Table 2-2. To evaluate the shape similarity, in this study metrics for all the indices are firstly computed and then representative metrics are selected. “Tangent Space Representation” is not implemented in this study, because experiments show that it is sensitive to details on the boundaries; it is also complicated to compute with inner rings of polygons.

Finally, for the Positional accuracy evaluation, the distances from corresponding check points between a reference and extracted building are measured. Such check points can be corner points or centroids: the distance at corresponding corner points is denoted as “dist(crn)”; the distance for corresponding centroids is referred to as “dist(ctr)”.

D. A Multi-criteria and Hierarchical Evaluation System

A multi-criteria evaluation system is built hierarchically with three levels in Figure 2-3: The per-building level at bottom describes metrics for each single building. The per-scene level in the middle describes each metric for a whole scene. The overall level
defines a summarized index. Currently most evaluation methods are developed at per-scene level and provide metrics for the entire study area; the proposed overall level aims to provide a single index for effective comparison of building extraction results. The per-building metrics can provide detailed information about the performance of a certain extraction method on individual buildings. It is useful for large scale mapping applications concerning the accuracy of each building. Detailed metrics provide valuable information for further manual editing and correction of each building from different perspectives.

The evaluation system, shown in Figure 2-3, is also divided into three different components as described in the Section 2.1.2. For the matched rate component, Comp, Corr and Q are computed at per-building level. Then Comp(I), Corr(I), Q(I) and Kappa(I) are calculated for all buildings in the whole image at per-scene level (e.g., Q(I) means the Quality for the image). For the shape similarity component, ratios of shape measures from the reference and extracted buildings are used. The image moment (Song & Haithcoat, 2005) similarity (distance metric) is calculated at the per-scene level. For the positional accuracy component, distances on corner points and polygon centroids are used. The corner points are referenced using ground surveyed points. When ground survey is unavailable, corresponding corner points from reference buildings provide an alternative.

Two analysis methods are employed to reduce the redundancy between metrics. Firstly, within shape similarity metrics, a principal component analysis (Jolliffe, 2002) is conducted to select representative metrics, denoted in Figure 2-3 as “Rep1”, “Rep2”, etc, which are used in the upper levels. Secondly, for all metrics at per-scene level, a correlation analysis is performed to remove highly correlated metrics between different components. Moreover, this system may be extended to evaluate 3D building extraction in the future. For example, for matched rate, the area on 2D plane is replaced by 3D volume of the building and/or the surface area; for shape similarity, building’s projection on the three dimensional planes can replace the current 2D plane for metrics’ computation; for positional accuracy, distance can be directly used in 3D space.
Figure 2-3. A multi-criteria and hierarchical evaluation system for building extraction. The system consists of three components (see columns): matched rate, shape similarity and positional accuracy; it also has three levels (rows), at per-building, per-scene, and overall levels, respectively.
To generate a final summarized index, different weights for the three components may be set by users, in accordance with specific application objectives. Although weights for the three components are subjective, an integrated and summarized index is easier to use for comparison purposes. The summarized index \((\text{Sum}_\text{Idx})\) can be computed as follows:

\[
\text{Sum}_\text{Idx} = \text{Matched_rate} \times w_1 + \text{Shape_similarity} \times w_2 + \text{Positional_accuracy} \times w_3
\]

where \(w_i (i=1,2,3)\) is the weight.

2.2 Experiments for building extraction methods

2.2.1 Study Area and Data Sources

The study area is located on the campus of the University of Western Ontario (UWO) in London, Canada (43° 0'34.55"N, 81°16'25.44"W), and its surrounding residential areas, as shown in Figure 2-4. The study area includes various land cover types: impervious surface, trees, rivers and bare grounds. Building footprint sizes also vary in the study area, according to building function. There are Tall institutional, and residential buildings on the campus of UWO and low residential buildings nearby.

LiDAR data and aerial Colour-Infrared (CIR) imagery were collected for the study area. The LiDAR data collected on 20 May, 2006, has an original density of 0.8-0.9 points/m²; the imagery was re-sampled and interpolated to a 1m resolution raster image. The optical image is CIR imagery with 0.3 m spatial resolution captured in June 2008 using Vexcel UltraCamX, by the City of London for a tree cover mapping project (Lehrbass & Wang, 2012). The CIR images include the green, red and near-infrared bands. The reference building footprint vector file comes from the City of London (2006), which is identical to building blueprint maps provided by university physical plant over the campus area. Building corners can be validated with collected ground reference GPS points, which were surveyed in October, 2011 using a differential GPS with absolute position accuracy within 1 meter. The GPS points only cover the campus area because most residential buildings are not accessible for measuring. Any buildings that underwent external structural changes from LiDAR data collection to CIR image acquisition were manually
removed from the evaluation. For comparison and analysis purposes, the study area is divided into nine different regions with roughly equal numbers of buildings in each region. The roads and rivers are used as region borders. Regions 1 to 3 are mainly public service buildings, while Regions 4 to 9 are mainly residential buildings.
Figure 2-4. The study area with reference buildings, ground GPS points and sub-regions.

The background is the CIR imagery.
2.2.2 The Four Different Building Extraction Methods

Four different methods for building extraction in the study area have been conducted as shown in Figure 2-5. The first three methods use the original CIR image, LiDAR and their combination, respectively. The fourth method has the same data processing steps as the combined CIR/LIDAR methodology, but the input CIR image is corrected from relief-displacement (Wang et al., 2012). For details about the relief displacement correction, please refer to Lehrbass and Wang (2012) and Wang et al. (2012). The four methods are:

1. The CIR image method (denoted as “CIRimage”), which uses the spectral information alone and extracts buildings via object-based supervised classification. This is the most straightforward method to extract building footprint from optical imagery. More details about object-based image classification are given in Appendix B.

2. The LiDAR method (denoted as “LiDAR”), which relies on the normalized DSM (nDSM), together with the height and object’s shapes for extraction. To distinguish buildings from other high objects, their shape and relationship to surrounding high objects are also utilized for extraction. More details about LiDAR data property and building footprint extraction from LiDAR are provided in Appendix C.

3. The method that combines LiDAR data with the original CIR image before the relief displacement correction (denoted as “LiDAR+unCorr”), which exploits both height and colour information. The colour information from the CIR images and the height information from the LiDAR data are combined in the object-based algorithm to develop a rule set. A better extraction result is expected.

4. The method that combines LiDAR data with the CIR image after the relief displacement correction (denoted as “LiDAR+Corr”), which aims to investigate the effect of relief displacement on building extraction. In an ortho-rectified CIR images based on a digital terrain model, tall buildings are still misaligned due to relief displacement. The cross-track relief displacement can be predicted for all points on the ortho-image (Lehrbass & Wang, 2012; Wang et al., 2012).
Figure 2-5. The procedure of performance evaluation of the four methods for building footprint extraction.
With this method, the tall objects can be aligned between the LiDAR and the CIR images. The object-based rule set is the same as in the M3 “LiDAR+unCorr” method.

Object-based segmentation and classification is used for Methods (1), (3), and (4); a decision tree is developed for these methods. Method (2) is implemented using LiDAR Analyst, which is commercial software for LiDAR data processing and feature extraction. LiDAR Analyst exploits the strength of each returned signal, the surface shape, and environment context (i.e. slope), to determine the ground feature types hierarchically. The results from each of the four methods are exported as vector maps and are further refined. To facilitate the comparison process, each building in the reference map is given a joint ID to be connected with a building in each extracted result. An overlap operation between reference and the extracted result can automatically detect simple “one-to-one” relations; for buildings with “many-to-one” relations, topological clarification (Rutzinger et al., 2009) is carried out to detect and process topological mismatch. An extracted building covering more than one reference buildings is split, while two or more extracted buildings inside the same reference building are merged. For “one-to-zero” and “zero-to-one” relations between extracted and reference buildings, no shape similarity is calculated.

The Area and Centroid for each building in the extracted and reference maps are computed; together with the overlap Area, TP, TN, FP, and FN are calculated. The corner points of the extracted buildings are identified by matching the corresponding corner points to the GPS survey points. Shape similarity metrics on polygons are computed by commercial software and saved as attributes for each polygon. The image moment based distance is computed between the rasterized images of the extracted and reference buildings.
Figure 2-6. A subarea showing the four building extraction results. (a) The extracted result from Method 1 -CIR image (M1); (b) The results from Method 2 -LiDAR (M2); (c) The result from Method 3 -LiDAR and CIR image before relief correction (M3); (d) The result from Method 4 -LiDAR and CIR image after relief correction (M4).
2.2.3 Building Extraction Result Maps

From Figure 2-6, the method M1 with CIR image presents the least accurate result based on visual interpretation, as this method extracts all the buildings with random shapes. Another problem is that buildings are confused with parking lots. The M2 method with LiDAR data provided a much better result, however, M2 has some extraction errors along the rivers in the densely treed areas. This is because LiDAR does not have colour information and buildings can only be extracted based on height and shape. In densely tree covered areas and with shapes similar to buildings, errors will occur. From visual evaluation, the results for M3 and M4 look similar. Both of them have high accuracy with all buildings extracted and non-building objects removed. However, when looking into details, the extracted building boundaries in M4 are “cleaner” with less unnecessary vertexes. Further analysis will be discussed based on computed metrics in the Result and Discussion (Section 2.4).

2.3 Selection and summary of metrics at different levels

2.3.1 Redundancy Reduction in Shape Similarity Based on Principal Component Analysis

When comparing the reference and extracted buildings at per-building level, twelve possible shape similarity ratios are used as listed in Section 2.1.2. Based on the result from M4 (“LiDAR+Corr”), a Principal Component Analysis (PCA) is performed on these ratios and the first four Principal Components (PC) of the PCA are shown in Figure 2-7. More details about PCA principles are provided in Appendix D. The biplot in Figure 2-7 (a) shows the coefficients of the first two principal components (PC1 and PC2) and Figure 2-7 (b) displays the biplot for the third and fourth components (PC3 and PC4). The magnitude and sign of each shape metric's contribution to the principal components are shown as the length and direction of 2D vectors, while samples are shown as dots. For example, in the biplot Figure 2-7 (a), the ratio for main direction vector has coordinates (0.8, -0.42), because its linear correlation coefficient to PC 1 is 0.8 and that for PC 2 is -0.42.
Figure 2-7. The principal component analysis for twelve shape similarity metrics. (a) The biplot of the principal components 1 and 2 with samples. Each red dot represents a sample building. The blue lines are vectors to show the contribution of shape descriptors. The coordinates for a ratio metric indicates its contribution to components PC 1 and PC 2 respectively. The percent in the axis label indicates the amount of variance accounted for by that component. (b) The corresponding biplot for PC 3 and PC 4.
From Figure 2-7, PC 1 accounts for 44.7% of variance and the major contribution comes from *Direction*, *Asymmetry* and *Roundness*; PC 2 covers 18.7% of variance and it is contributed by the components of *Direction* and *Roundness*. PC 3, with 12.5% of variance, is mainly influenced by *Asymmetry* and *Direction*; and PC 4, with only 7.4% of variance, is related to *Area*, *Perimeter* and *Roundness*. The accumulated variance from the first four components is about 85% of the total variance. Experiments on the other three methods (M1 to M3) report similar results. Based on the above analysis, the *shape similarity* metric at *per-building* level is represented by the metrics that significantly contributed to the first four PC components, including *Area*, *Perimeter*, *Direction*, *Asymmetry* and *Roundness*.

2.3.2 The Linear Correlation between Positional Accuracy Metrics at Per-building Level.

Two metrics are computed for positional accuracy: distances of corner points, dis(crn), and distances of centroids, dis(ctr). To analyze the linear correlation between them, the extracted building result from M4 (“LiDAR+corr”) is evaluated as an example. 47 building corner points were ground surveyed by a GPS. These points are matched with the corresponding corners of the extracted buildings on image to calculate corner point positional errors. For the buildings which the corner points were collected, their building centroids are computed and the corresponding distance between the extracted and reference building centroids are calculated. The relationship of the two distance metrics are shown in the scatter plot (Figure 2-8) with the two distance axes. The linear correlation coefficient for these two distance metrics is 0.133. As such, the distance metrics are not significantly correlated at 0.05 level. These results indicate that the corner point’s positional change is not significantly related to the centroid positional change. Therefore, the distance metrics on both corner points and building centroids are used to evaluate the *positional accuracy*. 
Figure 2-8. The scatter plot of the corner point distances and the corresponding building centroid distances for buildings.
2.3.3 The Aggregation of Duplicate Metrics Based on the Correlation Analysis.

At the per-scene level, different metrics are derived from the metrics for individual buildings (such as the mean ratio for shape similarity) or computed directly from per-scene level (e.g., the kappa coefficient and the image moment similarity distance). There are often too many metrics for effective evaluation, and there are possible correlations between them. To further compare and integrate these metrics, they are grouped by nine different regions shown in Figure 2-4. Using the extracted result from M4 (“LiDAR+uncorr”), all metrics in each region are separately computed and the correlations between the metrics are reported in Table 2-3.

In Table 2-3, $Q$ and Kappa represent the matched rate and they have very high correlation (0.99). Considering that $Q$ is concisely defined and easy to compute, Kappa is discarded in the upper level of evaluation. Many ratios for shape similarity are correlated to each other, especially the ratio for perimeter, asymmetry, main direction, and roundness. The $r(Area)$ and $Img_mnt$ are not significantly correlated to any others at 0.05 level. Therefore, the four correlated ratios are combined as only one single metric with equal weight, called “$r(Other)$”. At the upper level, three metrics are used to represent the shape similarity: $r(Area)$, the combined $r(other)$, and the $Img_mnt$. Although correlation matrices can vary for different study cases, the strategy to assess correlations and congregate them into relatively independent metrics is the same.

After the selection of metrics at per-scene level, there are few metrics left. For matched rate, $Q$ is selected; for shape similarity, $r(Area)$, $r(Other)$, and $Img_mnt$ are selected; for positional accuracy both the distances of building centroids and corner points are selected.

2.3.4 Summary of Metrics to the Overall Level

In the case when multiple building extraction methods need to be compared, the selected metrics at per-scene level can be further summarized as a single index. This summary, however, is subjective because it allocates individual weights for the three groups of the metrics. But a single evaluation index is the straightforward way to compare various
extraction methods. This summary, though subjective, is valuable and feasible if the application purpose is clear and experiments are conducted before giving weights to each metric group.

Two issues need to be considered in the summary. First, weights need to be assigned. For an application concentrated on classification accuracy, higher weight should be assigned to the *matched rate* because it is the most important criteria used to evaluate the result; for a 3D re-construction application, the weight for *shape similarity* should increase, because *shape similarity* is useful in determining the appearance of buildings. Currently, positional accuracy is still not a major concern in many applications; usually, lower weight can be given to this component. However, as future applications have stricter demands, requiring more accurate extraction, the positional accuracy will also become important. In this study, the *matched rate* (represented by \( Q \)) is given a weight of 0.5; the *shape similarity* (represented by \( r(Area) \), \( r(Others) \), and \( Img\_mnt \) with the equal weight of 1/3) is given a weight of 0.3; the positional accuracy (represented by \( \text{dist}(crn) \) and \( \text{dist}(ctr) \) with the equal weight of 1/2) is given a weight of 0.2.

The second issue is the strategy to normalize various metrics. The most commonly used strategy is to rescale the distance metrics to a range of 0 and 1, as follows:

\[
x' = \frac{(x - \text{min}(X))}{(\text{max}(X) - \text{min}(X))}
\]  

where \( X \) is the vector of all samples and \( x \) is an individual sample. \( \text{min} \) and \( \text{max} \) are the minimum and maximum values in the samples, respectively. \( x' \) is the new normalized metric between 0 to 1. However, this method sets the worst extraction method with a new metric value of 0, while the best extracted method will receive a new metric value 1. This may not be fair for metrics with different value ranges. As a result, an improved normalization changes the range as follows:

\[
x' = B + (A - B) \times \frac{(x - \text{min}(X))}{(\text{max}(X) - \text{min}(X))}
\]

where \( A \) and \( B \) are introduced as the new range. \( A \) and \( B \) are the *minimum* and *maximum* values among all *ratio* metrics values at current level, respectively. That is to say, the range of 0 to 1 is replaced by the current range of ratio metric values.
Table 2-3. The correlations between the metrics at *per-scene* level

<table>
<thead>
<tr>
<th></th>
<th>Q</th>
<th>Kappa</th>
<th>r/Area</th>
<th>r(Peri)</th>
<th>r(Asym)</th>
<th>r(Dir)</th>
<th>r(Rnd)</th>
<th>Img_mnt</th>
<th>M(ctr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kappa</td>
<td></td>
<td>0.99</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r/Area</td>
<td>0.48</td>
<td>0.44</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r(Peri)</td>
<td>-0.68</td>
<td>-0.67</td>
<td>-0.47</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r(Asym)</td>
<td>0.88</td>
<td>0.87</td>
<td>0.56</td>
<td>-0.60</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r(Dir)</td>
<td>0.78</td>
<td>0.76</td>
<td>0.43</td>
<td>-0.96</td>
<td>0.71</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r(Rnd)</td>
<td>0.67</td>
<td>0.62</td>
<td>0.56</td>
<td>-0.44</td>
<td>0.77</td>
<td>0.56</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Img_mnt</td>
<td>0.20</td>
<td>0.27</td>
<td>-0.59</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.13</td>
<td>-0.06</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>M(ctr)</td>
<td>0.43</td>
<td>0.40</td>
<td>0.27</td>
<td>-0.68</td>
<td>0.29</td>
<td>0.70</td>
<td>0.45</td>
<td>0.07</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: the correlation coefficients in **bold** indicate that they are significant at 0.05 level.
For the distance metric, lower values correspond to better extraction performance. Equation (2-3) has reversed the distance metric so that the higher value represents higher accuracy and better performance. This is consistent with other accuracy measures. Finally, the overall summarized index can be calculated by Equation (2-1) to compare the performances of all methods.

2.4 Results and discussions

2.4.1 Performance Assessment by the Evaluation System

Based on the buildings extracted by the four methods introduced, different metrics are computed to compare and evaluate the extraction methods. The metrics for each level are computed according to the evaluation system described in Figure 2-3. At the per-building level, PCA is used to select representative metrics from shape similarity metrics. At the per-scene level, metrics from the three different components are further congregated as based on the correlation analysis. The non-ratio metrics (e.g., centroid distance) are normalized to make them comparable with the ratio metrics. The metrics at per-scene level are reported in Table 2-4. With example weights for a general application (match rate: 0.5, shape similarity: 0.3, positional accuracy: 0.2), an overall summarized index for the building extraction results is reported in Table 2-5.

From the perspective of matched rate described as Quality in Table 2-4, the accuracy from M1 to M3 is gradually improved. This improvement indicates that the method with only CIR image has the worst matched rate, while LiDAR can improve the extracted accuracy from about 0.4 to 0.7. The combination of LiDAR with uncorrected CIR images slightly improves the extraction accuracy. M3 and M4 have very similar accuracy on matched rate (0.79 and 0.80), while they out-performed M1 (0.41) and M2 (0.70). From the perspective of shape similarity and positional accuracy, the four methods show a similar trend. From M1 to M4 in Table 2-4, the ratio metrics increase from the first to last (0.69, 0.77, 0.79, to 0.86), while non-ratio metrics show largest distance for M1 and decrease from M1 to M4.
It is difficult to compare extraction methods in Table 2-4 because of too many metrics; therefore, a single metric for shape similarity in Table 2-5 provides straightforward information for users to differentiate extraction methods. In contrast to matched rate, one noticeable difference is shape similarity and positional accuracy metrics in Table 2-5 show clearly that M4 performs better than M3, namely 4% and 9% respectively (compared with 1% for the Matched Rate).

The summarized index indicates that the performances of the four methods are improved from M1 to M4 in Table 2-5, which is consistent with detailed metrics in Table 2-4 and image visual assessment in Figure 2-6. When visually inspected, the method using only aerial imagery (M1) appears to have performed less effectively than the other three methods; with LiDAR data (M2), the extraction accuracy improves from 0.41 of M1 to 0.7; while the combination of these two data (M3) further improved the performance to 0.79. The difference for methods with LiDAR and aerial imagery before (M3) and after (M4) relief correction can be identified in the summarized index, while such a subtle difference cannot be pointed out by merely matched rate. This conclusion is consistent with the visual evaluation of Figure 2-6.

2.4.2 The Comparison of Evaluation Systems: Traditional Method vs. the Proposed System.

The comparison of the four building extraction methods on the entire study site has demonstrated that the evaluation system can provide consistent assessment with human visual evaluation. But individual buildings in the overview maps of Figure 2-6 are small and difficult to perceive. A few example buildings are compared between M3 and M4 to further analyze the performance of the evaluation method.

To analyze the performance of the proposed evaluation system, traditional methods based on the classification accuracy are used to compare with the proposed evaluation system. As introduced in Section 2.1.2, such traditional methods use matched rate metrics such as Completeness, Correctness, Quality, etc. Considering Quality is a popular metric for
evaluation and it has been applied in recent studies (Awrangjeb et al., 2010; Rutzinger et al., 2009). *Quality* is used as the traditional evaluation method to compare with the proposed evaluation system in this study.
Table 2-4. The metrics at *per-scene* level after congregation

<table>
<thead>
<tr>
<th>Method</th>
<th>Quality</th>
<th>r/Area</th>
<th>r/Other</th>
<th>Img_mnt</th>
<th>dist (ctr)</th>
<th>dist (crn)</th>
<th>Stndzd* Img_mnt</th>
<th>Stndzd dist(ctr)</th>
<th>Stndzd dist(crn)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.41⁺</td>
<td>0.69</td>
<td>0.65</td>
<td>40.25</td>
<td>4.95</td>
<td>4.47</td>
<td>0.41</td>
<td>0.41</td>
<td>0.41</td>
</tr>
<tr>
<td>M2</td>
<td>0.70</td>
<td>0.77</td>
<td>0.80</td>
<td>3.16</td>
<td>2.20</td>
<td>2.38</td>
<td>0.84</td>
<td>0.75</td>
<td>0.73</td>
</tr>
<tr>
<td>M3</td>
<td>0.79</td>
<td>0.85</td>
<td>0.76</td>
<td>4.90</td>
<td>1.54</td>
<td>2.51</td>
<td>0.82</td>
<td>0.84</td>
<td>0.71</td>
</tr>
<tr>
<td>M4</td>
<td>0.80</td>
<td>0.86#</td>
<td>0.83</td>
<td>1.44</td>
<td>1.36</td>
<td>1.53</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Note: M1: CIR image; M2: LiDAR; M3: LiDAR+uncorrCIR; M4: LiDAR+ corrCIR.

* Stndzd : “standardized”; “⁺” : the minimum metric value; “#”:the maximum metric value.
Table 2-5. The summarized index at *overall* level based on the three components

<table>
<thead>
<tr>
<th>Method</th>
<th>Matched rate (w:50%)</th>
<th>Shape similarity (w:30%)</th>
<th>Positional accuracy (w:20%)</th>
<th>Summarized index</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.41</td>
<td>0.58</td>
<td>0.41</td>
<td>0.46</td>
</tr>
<tr>
<td>M2</td>
<td>0.70</td>
<td>0.80</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>M3</td>
<td>0.79</td>
<td>0.81</td>
<td>0.77</td>
<td>0.79</td>
</tr>
<tr>
<td>M4</td>
<td>0.80</td>
<td>0.85</td>
<td>0.86</td>
<td>0.82</td>
</tr>
</tbody>
</table>
Figure 2-9. Four examples of the extracted buildings by two extraction methods (M3 and M4).
Table 2-6. The corresponding metrics of the four example buildings.

<table>
<thead>
<tr>
<th>ID</th>
<th>Area (m²)</th>
<th>Traditional method (Quality)</th>
<th>Shape similarity*</th>
<th>Centre distance (m)</th>
<th>This study (overall)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M3</td>
<td>M4</td>
<td>diff&lt;sup&gt;+&lt;/sup&gt;</td>
<td>M3</td>
</tr>
<tr>
<td>(a)</td>
<td>4973</td>
<td>0.88</td>
<td>0.93</td>
<td>&lt;strong&gt;0.05&lt;/strong&gt;</td>
<td>0.66</td>
</tr>
<tr>
<td>(b)</td>
<td>6069</td>
<td>0.92</td>
<td>0.96</td>
<td>&lt;strong&gt;0.04&lt;/strong&gt;</td>
<td>0.73</td>
</tr>
<tr>
<td>(c)</td>
<td>227</td>
<td>0.79</td>
<td>0.79</td>
<td>&lt;strong&gt;0.00&lt;/strong&gt;</td>
<td>0.62&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>(d)</td>
<td>881</td>
<td>0.83</td>
<td>0.83</td>
<td>&lt;strong&gt;-0.01&lt;/strong&gt;</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Note: +the minimum metric value; #the maximum metric value; ^ diff=M4-M3; *shape similarity at per-building level has no image moment distance, which has been replaced by perimeter ratios. **Bold** highlights the difference for comparison.
Four sample buildings extracted from M3 and M4 are analyzed. In Figure 2-9, extraction results from the two methods are overlaid on the reference buildings. Extracted buildings from the two methods are similar to each other, except that the results from M4 have cleaner boundaries with less noise and irrelevant parts. The improvement from M3 to M4 is mainly on the image segmentation stage, where M3 will generate more fragmented buildings because vertical LiDAR data sometimes conflict with the tilted buildings on the aerial image before relief displacement correction.

In Figure 2-9, human visual evaluation can easily identify that building boundaries from M3 are not well matched with the reference buildings. The traditional method with Quality, however, is not sensitive to the fluctuation on the border; the Quality difference between M3 and M4 is less than 0.05 in Table 2-6. In contrast, shape similarity metrics can detect this difference, with a reported 15% to 30% improvement from M3 to M4. Case (c) in Figure 2-9 is an extreme case, where Quality is coincidently the same between M3 and M4 (both 0.79). The extracted building from M3 is obviously less accurate than that of M4 compared with the reference building footprint in (c). It is the other components (shape similarity and positional accuracy) that can distinguish the two results. The overall summarized index for these buildings can differentiate the extraction accuracy that cannot be separated by the traditional method. Compared with the negligible Q difference between M3 and M4, the proposed evaluation system shows major differences in shape similarity and positional accuracy, with 4% and 9% respectively compared with 1% for traditional matched rate.

It can be noticed that the performance of metrics at per-building level in Table 2-6 is better than at per-scene level in Table 2-4. That is because per-scene level needs to consider “zero-to-one” and “one-to-zero” cases. The buildings that fail to be matched will decrease the overall accuracy.
2.4.3 Discussion of Building Types via the Multi-level System.

The multi-criteria and hierarchical evaluation system with three levels provides a framework of metrics. Such a multi-level system is valuable for tracing metrics from top to bottom. For example, a low overall index can be searched downward in order to identify the reason. The reason may be owing to the low shape similarity at per-scene level; the low shape similarity can be further investigated, it may be caused by low perimeter ratio at per-building level. Such an analysis provides feedback to further improve building extraction methods. Furthermore, metrics at the bottom level (per-building) can be sorted, grouped, or clustered, in order to discover the distribution of extraction errors. For instance, sorting extracted buildings according to a metric at per-building level can assist manual editing, primarily to modify the buildings with the lowest metric value. Moreover, grouping extracted buildings with a certain attribute can test systematic bias related to different types of buildings.

As an example, a building’s occupancy/function is investigated to analyze whether an extraction method is effective at a given building type. In Table 2-7, buildings are grouped as two types: public buildings and residential buildings; therefore, metrics at per-building level are grouped to re-calculate their per-scene metrics based on their types.

Metric values of public buildings are generally higher than that on the residential buildings. The public buildings are usually larger and higher (with less tree covering) than residential buildings; thus they are more easily detected. Two metrics significantly differ between large buildings and smaller ones: the distance between centroids and the ratio for perimeter. The centroid locations of larger buildings with complicated shapes can be affected when one corner is missed. The residential buildings are usually rectangles. Therefore, the centroid locations do not shift much between the extracted and the reference buildings. This analysis is important to improve the advanced algorithm design and to better detect buildings in residential areas.
Table 2-7. The metric comparison between public buildings and residential buildings

<table>
<thead>
<tr>
<th></th>
<th>Size(m²)</th>
<th>Comp</th>
<th>Corr</th>
<th>Q</th>
<th>kappa</th>
<th>dist(ctr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pub*</td>
<td>1420</td>
<td>0.94</td>
<td>0.89</td>
<td>0.84</td>
<td>0.90</td>
<td>1.48</td>
</tr>
<tr>
<td>Res#</td>
<td>150</td>
<td>0.74</td>
<td>0.87</td>
<td>0.67</td>
<td>0.79</td>
<td>1.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>r(Area)</th>
<th>r(Peri)</th>
<th>r(Asym)</th>
<th>r(Dir)</th>
<th>r(Rnd)</th>
<th>log_mnt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pub</td>
<td>0.89</td>
<td>0.77</td>
<td>0.87</td>
<td>0.91</td>
<td>0.83</td>
<td>0.17e-3</td>
</tr>
<tr>
<td>Res</td>
<td>0.85</td>
<td>0.85</td>
<td>0.78</td>
<td>0.79</td>
<td>0.79</td>
<td>0.39e-3</td>
</tr>
</tbody>
</table>

Note: * Regions 1-3 in the study area are grouped as public buildings (Pub); # regions 4-9 are grouped as residential buildings (Res).
2.5 Conclusion
This study proposes a multi-criteria and hierarchical evaluation system for building footprint extraction methods. After a review of the current methods and strategies used in building extraction evaluation, the problems of current evaluation methods are summarized as: overestimating the importance of matched rate, incomplete description of shape similarity, overlooking the redundancy between metrics, and lack of an overall index for comparison purpose. Consequently, a multi-criteria evaluation system represented by three components from different perspectives is proposed. These components are: (a) the matched rate that describes the proportion of the building area successfully extracted when compared with the reference data, (b) the shape similarity that describes to what extent an extracted building is similar to its reference data, and (c) the positional accuracy which evaluates the geometric distance between extracted and reference buildings.

The proposed evaluation system is also stratified into three levels: the per-building level, per-scene level, and the overall level. The per-building level evaluates individual buildings with different metrics; it can provide detailed information for each building and it is valuable for manual editing after extraction. The per-scene level treats each scene as a unit for evaluation and different metrics are provided to assess the whole scene. It is the major level for most evaluations. The overall level is used in an attempt to summarize different metrics as a single index in order to perform an integrated comparison of different building extraction methods. Although the overall level is summarized by setting weights subjectively, it provides a concise manner for comparison. Future work may discuss how to obtain reasonable weights based on reliable survey among users or experiments.

The proposed evaluation system is tested by four building extraction methods. The four methods use different data sources, including the original CIR images, LiDAR data and their combination. The last method combines LiDAR data with CIR image after relief
displacement correction. These four methods are compared. The proposed evaluation system demonstrated its superiority to traditional evaluation methods, especially when traditional classification accuracy cannot distinguish the performance of two extraction methods, while the proposed system can identify the best performer very well. Finally, this multi-level system permits analysis of the impact of building types on evaluation methods.

In summary, this study explores the evaluation methods to assess building extraction accuracy from remotely sensed imagery. Specifically, (a) to complete the description about the shape similarity metrics, twelve shape metrics are developed at the per-building level and principal component analysis is conducted to remove redundancy between metrics; (b) to remove correlations between different metrics, correlation analysis at the per-scene level are utilized to select and congregate metrics. Congregated metrics simplifies the comparison at the per-scene level with less relatively independent metrics. (c) To make the comparison of methods straightforward, an overall index is proposed by using the weighted linear combination of the three evaluation components. This study also provides a way to integrate ratio and distance metrics by rescale the latter. (d) To test the proposed evaluation system, four different building extraction methods are evaluated by the proposed system. Evaluation method from this study performs more consistently under visual inspection than traditional evaluation methods, since other dimensions (i.e., shape similarity) also play a role in evaluation outcomes.

References


Chapter 3: A Stereo Image Matching Method for Building Roof Elevation Estimation Assisted by Building Footprints*

3.1 Introduction
In urban areas, a digital surface model (DSM) provides important data for applications such as urban feature extraction, three-dimensional (3D) reconstruction, urban planning, flood prediction, and visualization. DSMs can be derived from various data including Light Detection and Ranging (LiDAR) data (Lloyd & Atkinson, 2002; Ma, 2005), synthetic aperture radar (SAR) images (Toutin, 2004b), and stereo pairs of optical images (Altmaier & Kany, 2002; Toutin, 2004a). Among these data, stereo pairs of optical images have been extensively used to generate DSMs because of their high spatial resolution, low cost, and large-area data coverage. Image matching is the most important method to generate DSM from stereo images (Birchfield & Tomasi, 1998; Foerstner, 1982; Hirschmuller, 2008).

Buildings are major objects in urban areas, and as such, the accuracy of building rooftop elevations in a DSM is critical for many applications. Current stereo image matching methods, however, report a lower DSM accuracy in dense urban areas than open areas (Alobeid et al., 2010; Capaldo et al., 2012). In dense urban areas, buildings impede successful image matching, resulting in low accuracy. Impeded factors include occlusions, shadows, and abrupt elevation change on building edges; in addition, extremely tall buildings and their exceptionally large image parallax also affect matching. Extremely Tall buildings is a relatively expression, which defines about the top 5% percent of all buildings a large area. This objective of this chapter is to develop a new image matching method to improve the DSM accuracy over building rooftops.

3.2 Literature review
Stereo image matching is categorized as feature-based matching and intensity-based matching (Gruen, 2012). In feature-based matching, features in an image such as corners and edges are first extracted and then matched (Kim et al., 2001; Suveg & Vosselman, 2004). In intensity-based matching, cost functions are used to measure the similarity between two images according to their image intensity. Commonly used cost functions include normalized cross-correlation (NCC) and least squares (Gruen, 1985), etc. Feature-based matching is challenging for automatic reconstruction, due to the difficulty in 2D feature extraction, expression, and matching. For example, broken and discontinuous edges are difficult to handle in matching (Zhang & Gruen, 2006). In contrast, a DSM extraction from intensity-based matching is a well-developed technique at local, global or semi-global levels (Gruen, 2012) and is widely used in current software.

In image matching, DSM error in urban areas is reported to be three times higher than the error in open areas (Capaldo et al., 2012). Buildings are the primary objects in urban areas and one of the DSM error sources. Specifically in intensity-based matching, local matching methods using window-based matching techniques fail to match left and right image pixels around buildings. This failure will lead to DSM error. In global (Boykov et al., 2001) and semi-global (Hirschmuller, 2008) stereo image matching, smooth constrain is usually used to remove noise. Smooth constrain, however, affects the discontinuity of ground features including building edges in the DSM. Furthermore, extremely tall buildings with a large parallax in stereo images increase the chance of mismatch. Equally, the perspective projection distortion of tall buildings makes successful matching difficult, resulting in an incomplete shape of buildings in the DSM.

To reduce the DSM error on building rooftops, geometrical constraints and prior knowledge (Gruen, 2012) can be used to limit the matching radius and eliminate irrelevant matching candidates. For example, epipolar geometry (Lillesand et al., 2008) is a widely used geometrical constraint that causes the matching conduct to occur only in one direction. Multi-view image constraint (Bulatov et al., 2011; Zhang & Gruen, 2006)
utilizes three or more images to provide redundant information for more accurate matching than one stereo image pair. Other Geographic Information System (GIS) database layers and maps can also be used as prior knowledge to assist matching because these data provide information about the location and shape of ground features. For example, building footprint maps are commonly available in urban areas from independent sources (e.g., cadastral maps) or remotely sensed images (Aldred & Wang, 2011). In an application concerned with the accuracy of elevation rooftop but not the entire area, it is reasonable to concentrate the image matching specifically on buildings.

Given a building footprint map, in Tack et al. (2012) an indirect method is adapted to improve the DSM accuracy. A building footprint is used to clip and refine the extracted DSM, and the average elevation values over building rooftops are replaced by more accurate elevations. Under the assumption that buildings are prismatic shaped structures with flat roofs, an appropriate building elevation is chosen from its elevation statistic variants such as mean, median, and maximum. This is a robust approach and easy to operate for a large area; nevertheless, the assumption that the building is flat is not always true and the selection of statistic variants for the building demands known building elevations. Rather than using a DSM after matching, a method that improves the building elevation directly from the matching stage is expected to provide more accurate elevation on buildings.

In summary, it is difficult for current stereo image matching methods and software to generate accurate DSMs in dense urban areas especially on building rooftops, considering shadows and occlusions, complicated scenes, and extremely tall buildings. To overcome the problem of inaccurate DSMs, extra data are required to narrow down matching candidates, remove mismatches caused by complicated scenes, and strengthen the matching process. A building footprint map is a popular and commonly available source of extra data that meets such requirements. However, to date, no published method uses building footprints directly in the matching stage to constrain and improve the matching. As a result, the objective of this study was to propose a new matching method that
integrates a building footprint map directly into the stereo image matching process, to improve the DSM accuracy over building rooftops.

3.3 Method: stereo matching assisted by building footprints
Stereo image matching assisted by building footprint maps exploits the prior knowledge provided by building footprints. Above all, a building footprint in the map can provide an initial location for the building. A building footprint overlaid on the image as an object makes it possible to match at the object level rather than the traditional pixel level; this is valuable in case pixel-based matching is difficult in complicated scenes. Furthermore, building footprints can validate the matching process through indirect clues such as edges on a building’s boundaries and the colour contrast inside and outside building boundaries in images. All of these rules are employed in the matching process.

The matching method proposed in this study follows a hierarchical order. The approximate location of a building is searched based on its footprint; this step is called “rooftop identification”. When determining the approximate location of a building, a building footprint provides the precise shape of a building to constrain the matching between the left and right stereo images; this step is called “left-right matching”. The final step adds elevation details on building rooftops, called “detailed matching”. This step has adapted existing pixel-based matching algorithms.

3.3.1 Rooftop identification
In “rooftop identification”, the complicated position of a building in imagery is identified using building footprints as a template. The combination of feature and intensity information for matching is reported to improve the matching accuracy (Wu et al., 2012). For a building’s intensity (colour) information, however, no appropriate parameter is available to define the intensity difference between a building and its background. A rooftop does not necessarily have a constant colour, as roof colour can change over time due to construction material, surface painting, and shadows; equally, there is no assurance that the rooftop’s colour should be different from the background. In contrast, edge features in images provide a concise representation of features in the image and they
can be less sensitive to reflectance characteristics (Gruen, 2012); therefore, extracted edge images were used in this step.

To identify the building rooftop in the stereo images, the building footprint template moves on edge images to search for the optimum matching. As shown in Figure 3-1, matching rate is the ratio of the intersected edge length to the total edge length of the template, counted as:

$$r(\text{template}) = \frac{|E \cap T|}{|T|}$$  \hspace{1cm} (3-1)

where $r$ is the matching rate, $T$ is the set of edge pixels in the template, $E$ is the set of edge pixels in the extracted edge image, $\cap$ is the intersection between two sets, $|.|$ is the cardinality (number of elements) of a set. The template moves around the initial location given by the footprint; matching rates at different positions form a matrix as shown in Figure 3-1(d). The position that corresponds to the maximum matching rate $r$ is the optimum position. For a pair of stereo images, each edge image matches with the building footprint template separately.
Figure 3-1. A conceptual example of rooftop identification. (a) A building footprint map with four buildings. (b) The building in an epipolar image corresponds to footprint D. (c) The footprint D moves to match with the edges derived from a stereo image. (d) The cross-section of the matching rate matrix in the X direction.
3.3.2 Left-right matching

With the approximate location for a building in epipolar images, the left-right matching aims to determine the best parallax for the entire building. Colour and edge information are both employed in the left-right matching. Similar to the rooftop identification step, the edge matching is conducted on left and right edge images without a template. An incentive is given to the correct matching on the boundary template. The edge matching rate for left-right matching as follows:

\[
r(\text{edge}) = \frac{|E_l \cap E_r| + |E_l \cap E_r \cap T|}{|E_l|}
\]  \hspace{1cm} (3-2)

, where \(E_l\), \(E_r\), and \(T\) represent the edge pixel sets in the left edge image, right edge image, and the footprint template respectively. Generally, the left image is assumed to be the near-nadir view image with less projective distortion.

For the colour information matching, normalized cross correlation (NCC) (Zhang & Gruen, 2006) is adapted to represent the correlation between the left and right subsets of the stereo images.

\[
r(\text{intensity}) = \text{NCC} = \frac{1}{n} \sum_{s \in W} \frac{(I_l(s) - \bar{I}_l) \times (I_r(s) - \bar{I}_r)}{\sigma_l \times \sigma_r}
\]  \hspace{1cm} (3-3)

where \(\bar{I} = \frac{1}{n} \sum_{s \in W} I(s)\) , \(\sigma = \sqrt{\frac{1}{n} \sum_{s \in W} (I(s) - \bar{I})^2}\) \(I_l\) and \(I_r\) denote the sub-images of left and right intensity image, \(W\) is the template window, \(s\) is a pixel in the window, and \(n\) is the number of pixels in the window. For a multiband image, an average NCC of all bands is utilized. The overall matching rate can be summed with given weights and is defined as:

\[
r = w_1 \times r(\text{edge}) + w_2 \times r(\text{intensity})
\]  \hspace{1cm} (3-4)

The optimum left-right matching corresponds to the maximum overall matching rate. A threshold \(T\) for the maximum overall matching rate is set for a successful matching. After the experiments in this study, a reasonable threshold \(T\) can be the mean value for best matching rates of all buildings minus two or three times their standard deviation, depending on quality of the detected edges and the complexity of the scene.
To efficiently match buildings and to reduce the chance of mismatch in large areas, a pyramid for left-right matching is carried out. Buildings have different base elevations and heights, and thus different parallaxes, whereas most buildings have a small parallax in epipolar images. The pyramid is designed to match buildings iteratively from a smaller to larger parallax. As illustrated in Figure 3-2, given a maximum possible parallax \((\text{maxD})\) for a study site, three different levels of matching are conducted. For the first round, only successful matching with the optimum parallax less than a 1/4 of \(\text{maxD}\) is accepted; all other matching results are ignored. The successfully matched buildings are masked and removed from the edge and intensity images. The second round increases the acceptable parallax to 1/2 of \(\text{maxD}\) and the last round uses \(\text{maxD}\). Buildings not matched after the three rounds are discarded. The three levels of matching can effectively reduce the chance of mismatch for low buildings and help to successfully match extremely tall buildings.

### 3.3.3 Detailed matching

After left-right matching, the accurate position and the average height of each building are determined. To further create rooftop elevation details, a pixel-based image matching is required. Considering its computation efficiency and sub-pixel matching accuracy, the semi-global matching (SGM) algorithm (Hirschmuller, 2008) is adapted to match subset images for detailed parallax on rooftops. The SGM combines the concepts of global and local methods for pixel-wise matching using mutual information between images. Sub-images are cropped from the left and right epipolar images and a base parallax for those two subsets is determined by the left-right matching. As a result, the overall parallax of each pixel is:

\[
D(s) = D_b + D_d(s) \tag{3-5}
\]

where \(s\) is a pixel in the subset, \(D\) is the total parallax, \(D_b\) is the base parallax calculated from left-right image matching, and \(D_d\) is the detailed image parallax over the building rooftop.
Figure 3-2. The pyramid for the left-right matching with three levels. “Rnd” mean a round of matching. “maxD” is the maximum possible disparity set by users.
3.4 DSM generation based on the proposed matching method

3.4.1 Study area and data description

The study area covers two city blocks in Beijing, China, as shown in Figure 3-3. Building density in the study area is high and buildings are constructed for a variety of purposes, such as residential buildings with a long narrow shape, university halls with bright roofs, and tall commercial office buildings along streets.

The stereo satellite image pair is GeoEye-1 imagery, captured at 11:00AM local time, 27 October 2010, with one 0.5m panchromatic band and four 2m multi-spectral bands (blue, green, red, and near-infrared). The sun azimuth angle was 164.6° and sun elevation angle was 36°. For the first image, the sensor azimuth angle was 204.6° and its elevation angle was 81.7°. This image is referred to as “left image” in the experiment. For the other image, the sensor azimuth angle was 10.5° and its elevation angle was 61.7°; this image is referred to as “right image”. The building footprint map was digitized from independent ortho-photos. The surveyed rooftop elevation points were used as ground truth. As most rooftops were not accessible, the combination of a differential GPS measurement and a laser range finder were used, where GPS measured the ground elevation and the range finder measured the building height. The error of elevation in the GPS points was within 1m and the range finder has an accuracy of 30cm. As shown in Figure 3-3, the dots are GPS points and crosses are range finder measured roof height points. The process of GPS and range finder data collection procedure and a record sheet example are provided in Appendix E. Three GPS points were used as GCP for image orientation and epipolar images generation.
Figure 3-3. The satellite image of the study area overlaid by the ground surveyed points
3.4.2 The framework for DSM generation

With the stereo images and building footprints, a framework is designed to extract DSM from the given stereo images, as shown in Figure 3-4. Figure 3-4 consists of three columns: the left column lists the overall steps for DSM generation; the middle column lists the detailed steps about image matching; the right column describes the edge detection techniques used in this study. In the left column, starting with a stereo image pair, the rational polynomial coefficients (RPC) define the relationship between image and ground; additionally, the ground control points (GCPs) are used for image orientation. Integration of the GCPs into the RPC (Tao & Hu, 2001) is reported to improve the accuracy of the orientation model. RPC, GCPs, and user defined tie-points establish the connection between image space and ground space. After image orientation, a common step is to rectify images according to the epipolar geometry (Szeliski, 2010). In epipolar images, the image parallax only happens X direction. After the stereo image is matched based on the method proposed in this study, parallaxes of buildings are computed. Parallaxes can be converted to ground distance in the real-world given the focal length and baseline of the images (Lillesand et al., 2008); the distance map is further wrapped toward a grid of ground coordinate system.

The matching stage process is the point on which the proposed matching method differs from previous methods. In this study, the building footprint is integrated into the matching process, as elaborated in the middle column. Before image matching, the preprocessing (mostly edge detection) is further described in the right column. After preprocessing, the proposed image matching method can be divided into building rooftop identification, left-right matching, and detailed matching, as described in Section 3.3.
Figure 3-4. The flowchart for DSM generation with the proposed image matching method
3.4.3 Edge detection

As the edge detection in preprocessing is a critical step because the quality of detected edge affects further matching. In the right column of Figure 3-4, the epipolar images are first pan-sharpened if both panchromatic and multiband images are available; edges are detected from the pan-sharpened images. Comparatively, traditional edge detection methods work on a single band, two band selection strategies: as such, the colour space transformation from RGB to Lab colour space (Hunter, 1948) and principle component analysis are investigated. Experiments demonstrate that the two methods generate similar edge images. Lab colour space transformation is applied in this study. Colour space transformation converts a pseudo-colour image (R, near-infrared; G, red; and B, green) into a Lab colour space. The L component is used as the band for edge detection.

To extract accurate building edges from images, different edge detection methods are compared. From initial experiments, the Canny operator out-performs other traditional edge detection operators and generates the best edge image for buildings. Colour edge detection is an alternative to the traditional methods for edge detection, considering its ability to detect edges at different colours but same luminance. For instance, the Compass edge detection in (Ruzon & Tomasi, 2001) is one colour edge detection method. Furthermore, the global Probability-of-Boundary (gPb) contour detector (Arbelaez et al., 2011) that out-performs most of other contour detection methods and close to human recognized boundaries, also involves in the comparison. Figure 3-5 provides a comparison of these three edge detection methods using a single sample image.
Figure 3-5. The comparison of edge detection methods. (a) The L component image in the Lab colour space. (b) The Compass edge detection. (c) The gPb image segmentation and edge detection. (d) Canny edge detection.
Through visual inspection, the Canny operator appears to perform better on building edges, in comparison to the other two operators. As a result, the Canny operator is selected. Shadows in the image are segmented, colour stretched, and input in an extra round of edge detection; however, the edges extracted from shadow areas are mainly noise. As only the edges of buildings are interested in this study, edges for vegetation can be safely removed. A Normalized Difference Vegetation Index (NDVI) derived from the same image is used to remove edges of vegetation surrounding buildings, using a threshold of NDVI (>0.6). Finally, trivial edges are removed using a minimum edge length threshold (<5 pixels in this study).

3.4.4 The stereo image matching assisted by building footprints

After the epipolar images and edge images are both ready, the image matching process assisted by building footprints is conducted on each building. For a given building in the footprint map, matching with the edge images aims to find the approximate location of the building in left and right epipolar images. The searching happens on the X direction of epipolar images, using a few extra pixels in width as tolerance. A threshold for successful matching rate was set: if the maximum matching rates for a building is lower than the mean value of all buildings’ maximum matching subtracting two times their standard deviations, this building failed to be accurately matched. In case that the a matching rate matrix has multi-peaks and the same peak values, the peak that is the closest to the initial position given by the template is counted as the correct match. Buildings that fail at the template matching stage will not be matched further.
Figure 3-6. The left-right matching rates for a sample building, including rates for edge matching, intensity matching, and overall matching.
Figure 3-7. An example of the stereo image matching process: (a) a picture of the building; (b-c) the left and right epipolar image subsets; (d-e) detected edges for left and right images; and (f) 3D building model after matching. The scale bar is valid for (b)-(e).
In the left-right matching, with a sub-image fetched from the left image, a sliding window moves on the X direction of the right image to match at each location. Both edge and intensity images are used and matching rates at each position are recorded. The edge matching rate is computed according to Equation (3-2); for intensity matching, the NCC is computed as the average of four different image bands, as in Equation (3-3). The overall matching rate combines the edge and intensity matching rates as described in Equation (3-4). Experiments demonstrate that the edge matching is more stable in the optimal matching (see Figure 3-6), while the contribution of intensity matching is minor. In this study, the weight for the edge and intensity is 0.8 and 0.2 respectively. An example of the matching process is given in Figure 3-6 and Figure 3-7.

As shown Figure 3-7, a university library located at southeast corner of the study side is selected as an example. The building footprint has been used to locate the approximate position of the building in the left and right GeoEye-1 epipolar imagery as given in Figure 3-7 (b) and (c). Edges are detected and refined and are shown in (d) and (e). Edge and intensity images are matched, with the left sub-image as the reference and a window sliding in the right image. The result of matching rate is given in Figure 3-6. In Figure 3-6, the edge matching rate is a curve with only one major peak, whereas the intensity matching has multiple peaks. That is because shadows conceal intensity difference on ground, create low intensity regions in images, and lead to a high intensity matching rate for those areas. The summed overall matching rate is a compromise for the two components; given the large weight (0.8) to edge matching, the overall rate curve follows the trend of the edge matching rate but slightly changes by intensity matching rates. The optimum matching has high matching rates for both edge and intensity matching.

After the left-right matching, a detailed matching is conducted on the sub-images determined using SGM. The SGM algorithm is implemented in openCV (Bradski & Pisarevsky, 2000) which replaces the mutual information cost function as a simpler Birchfield-Tomasi sub-pixel metric (Birchfield & Tomasi, 1998). P1 and P2 for SGM are set as 32 and 96 respectively, after many experiments. Figure 3-7 (f) is the elevation of the building derived from this study, it has an overall flat rooftop but there are some extra parts that are extracted by detailed matching. The details of the building show the two
extra buildings and the radar antenna on the rooftop, interpreted according to Figure 3-7 (a).

3.5 Accuracy assessment and Discussion

3.5.1 The comparison with commercial software

To analyze the accuracy of the matched building elevation, two popular commercial software modules (ENVI 4.8 and OrthoEngine in PCI v10.1) are compared with the proposed method in this study. The DSM for the same study area is generated using the same stereo images, GCP points, and tie points. During DSM extraction, PCI set the “extraction detail” as “high”, whereas in ENVI the “terrain details” is set as “maximum” and the “terrain relief” is set as “low (flat)”. Twenty-five surveyed rooftop points are used as reference points to evaluate the accuracy of the stereo image matching methods. The elevations of these points are extracted from the resultant DSMs from the three different DSM generation methods respectively; the elevation error is calculated based on the differences of the extracted elevations from the DSM and the ground surveyed elevations. The mean absolute error (MAE) and standard deviation (STD) for rooftop elevations are reported in Table 3-1. To analyze the effect on tall buildings with the DSM generation, roof points with elevation higher than 70 m are further analyzed and the detailed elevation errors are recorded in Table 3-2.

Table 3-1 demonstrates that DSMs derived from ENVI and PCI have elevation errors on rooftops three times or more than the errors in the proposed method. The reported error is higher than the general level of elevation errors derived from stereo images with same spatial resolution (Capaldo et al., 2012; Eckert & Hollands, 2010; Sirmacek et al., 2012), as the samples investigated in Table 3-1 are only on the rooftop; other studies use samples distributed in whole study areas. Although the surveyed reference rooftop elevations have accuracy within 1.5m, the proposed method generates rooftop elevations with accuracy at about 3m level in this study. Furthermore, the errors for commercial software are doubled for high rooftop elevations, whereas the elevation accuracy of the proposed method is not affected by the building rooftop elevation. When looking into details in Table 3-2, the matching method used by ENVI fails on most of the tall buildings, considering the average ground elevation (about 45m) is included in the
building’s elevation. PCI Orthoengine performs better on tall buildings but there are still some buildings mistakenly matched. The method proposed in this study improves the performance (with elevation error drops 80% or more) on extremely tall buildings and the matching method does not affect by the elevation of the buildings.
Table 3-1. The rooftop elevation errors for different stereo image matching methods

<table>
<thead>
<tr>
<th>DSM on rooftop</th>
<th>Err_ENVI</th>
<th>Err_PCI</th>
<th>Err_This study</th>
</tr>
</thead>
<tbody>
<tr>
<td>All samples (n=25)</td>
<td>MAE(m)</td>
<td>18.07</td>
<td>10.15</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>16.27</td>
<td>12.79</td>
</tr>
<tr>
<td>High elevation* samples (n=12)</td>
<td>MAE(m)</td>
<td>33.89</td>
<td>15.63</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>7.61</td>
<td>15.61</td>
</tr>
</tbody>
</table>

* High elevation denotes rooftop elevation greater than 70m.
Table 3-2. Detailed elevation errors for building roof points with elevation larger than 70m

<table>
<thead>
<tr>
<th>Point ID</th>
<th>Elevation (m)</th>
<th>MAE_ENVI (m)</th>
<th>MAE_PCI (m)</th>
<th>MAE_This study (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>73.92</td>
<td>22.39</td>
<td>8.93</td>
<td>4.31</td>
</tr>
<tr>
<td>13</td>
<td>75.89</td>
<td>27.05</td>
<td>8.86</td>
<td>1.58</td>
</tr>
<tr>
<td>8</td>
<td>78.60</td>
<td>29.81</td>
<td>3.59</td>
<td>2.33</td>
</tr>
<tr>
<td>2</td>
<td>82.44</td>
<td>34.30</td>
<td>4.73</td>
<td>0.17</td>
</tr>
<tr>
<td>23</td>
<td>83.09</td>
<td>26.91</td>
<td>19.04</td>
<td>1.64</td>
</tr>
<tr>
<td>1</td>
<td>83.71</td>
<td>33.64</td>
<td>2.72</td>
<td>1.44</td>
</tr>
<tr>
<td>24</td>
<td>85.93</td>
<td>28.62</td>
<td>8.94</td>
<td>4.76</td>
</tr>
<tr>
<td>16</td>
<td>86.85</td>
<td>34.14</td>
<td>13.48</td>
<td>0.43</td>
</tr>
<tr>
<td>4</td>
<td>87.86</td>
<td>37.48</td>
<td>19.11</td>
<td>3.94</td>
</tr>
<tr>
<td>25</td>
<td>91.55</td>
<td>36.91</td>
<td>1.60</td>
<td>6.81</td>
</tr>
<tr>
<td>15</td>
<td>97.79</td>
<td>44.67</td>
<td>47.64</td>
<td>6.40</td>
</tr>
<tr>
<td>20</td>
<td>102.59</td>
<td>50.72</td>
<td>48.93</td>
<td>3.48</td>
</tr>
</tbody>
</table>
3.5.2 The comparison and visualization of 3D buildings

With building’s rooftop elevation details, 3D buildings can be easily reconstructed under the assumption of flat ground and simple vertical walls. To visually compare different methods at the single building level, 3D buildings of the sample building are reconstructed and visualized as shown in Figure 3-8. The rooftop elevation from ENVI stereo image matching module has many errors because of stereo image matching mistakes for this tall building. This failure leads to elevation mistakes and an obviously wrong 3D building model. The 3D building from PCI OrthoEngine stereo image matching results in a rooftop elevation close to the actual rooftop elevation. The distribution of the elevation, however, will be fragmented with no clear roof shape if prior roof shape knowledge is unavailable. The majority of the roof elevation is underestimated with some obvious mistakes. The SGM is claimed to be an advanced method (Hirschmuller & Bucher, 2010) for stereo matching; however, to date there is no commercial software implementation. An approximation in OpenCV (openCV, 2013) reported in Figure 3-8(c) does not give satisfactory results. Compared with building footprint (the shadowed area at the bottom of Figure 3-8(c)), a notable portion of the building is missed; the rooftop is quite patchy but the pattern of roof objects is more exact than shown in (b). The last result from this study is based on SGM, after overall matching and determining average elevation. It provides the building with not only the overall shape of the building but also the elevation details.

By superimposing the improved elevations of building areas on DSM derived from image matching method or a flat background, a Digital building model (DBM) can be created. DBM describes the building structure, three-dimensional (3D) coordinates, and topologic relationship, etc (Zhou et al., 2005). It is an important data source for many applications such as true-orthorectification and visual volumetric representation.
Figure 3-8. Visualization of generated DSM and 3D buildings. (a) the 3D building model derived from ENVI matching model, with the rooftop location float on the top. (b) the corresponding result from PCI OrthoEngine. (c) the 3D building model generated using the SGM algorithm. (d) the 3D building model by the proposed method in this study which is stereo image matching constrained by building footprints.
3.6 Conclusion
This study investigates stereo image matching methods for buildings based on stereo satellite imagery and building footprints. Because of defects in the imaging process and the weakness of stereo image matching algorithms, elevation accuracies in dense urban areas are usually lower than those in open areas. A building footprint offers valuable data to constrain the matching process, but no current methods employ building footprints directly in the matching stage. To reduce DSM error over buildings, this Chapter proposes a novel stereo image matching method integrating building footprints into the matching process of stereo satellite images.

The major contribution of this study is the proposed matching method. A building footprint map not only provides the location of a building, but also provides prior knowledge about the complicated shape and size of the building in the image. Such information can narrow the range of matching candidates and reduce computational costs. Under the framework of current DSM extraction, a new stereo image matching method is designed to integrate building footprints. The designed matching method is divided into several steps. Before image matching, stereo images are preprocessed. The preprocessing extracts building edges, refines the edge maps by eliminating edges of vegetation, and cleans trivial edges. In image matching, a building footprint is first used as a template to identify the location of the corresponding rooftop in epipolar images. Second, left and right epipolar images are matched at the given building according to their edge and intensity similarity. Third, a detailed matching is conducted to refine elevation details on rooftops. A popular semi-global stereo image matching method is used for the detailed matching. A successful left-right matching demands high matching rates on both edge and intensity matching. To effectively match buildings of different heights, a pyramid of three-level matching is designed by increasing the threshold of the maximum acceptable image parallax.

In comparison with the DSM generated from other popular commercial software, the DSM created in this study demonstrates the superiority for building rooftop elevation extraction with a high accuracy. Experiments have been carried out with GeoEye-1 stereo
images in Beijing comparing the proposed method to two commercial stereo image matching modules (ENVI and PCI). Validation by field survey indicates that the proposed method has decreased the rooftop elevation error to one third or less than the current commercial software. Furthermore, whereas the commercial software doubles the error for tall building rooftop elevations, the proposed matching method keeps elevation accuracy for low and tall buildings consistent. The comparison between a direct SGM method and this study also indicates that the proposed method can generate a more accurate building than the SGM algorithm. The DSM, after refining building areas, can apply to image true-orthorectification, 3D building reconstruction, and other applications.

References


Chapter 4: An Elevation Difference Model for Building Height Extraction from Stereo-image-derived Digital Surface Model*

4.1 Introduction

Building height is an important data source for 3D building reconstruction, population density estimation, and natural disaster impact assessment (Brunner et al., 2010). Different approaches for building height extraction have been developed using three main data sources: Light detection and ranging (LiDAR) data, Synthetic aperture radar (SAR) imagery, and optical imagery.

The range measurement mechanism of LiDAR (Lillesand et al., 2008) provides a solid theoretical background for building height estimation. LiDAR data is not only used to estimate building’s height, but also to reconstruct 3D rooftops (Sohn et al., 2008). Building roof-top planar patches are typically segmented, clustered, and reconstructed with geometric limitation or other knowledge (Comber et al., 2011; Pu & Vosselman, 2009; Sampath & Shan, 2010). However, the prohibitive cost of LiDAR data and the restriction on flight plans in some countries limit its application in large urban areas.

SAR images have been used for building height estimation for two reasons: their all-weather reliability, and the capacity for high spatial resolution imaging in recently developed radar satellite sensors (such as the TerraSAR (Buckreuss et al., 2008)). Different strategies for building height estimation by SAR images include the double-bounce effect (Franceschetti et al., 2002), height hypothesis and simulated scene (Brunner et al., 2010) and the range size of layover and shadow together with view angles (Guida et al., 2010). Nevertheless, urban centres can be very difficult to interpret in a SAR image, considering the increase of multiple scattering in dense building areas (Guida et al., 2010). Most studies using SAR to derive building extraction are still limited

---

* A version of this chapter has been submitted and under 2nd round revision as “Chuiqing Zeng, Jinfei Wang, Wenfeng Zhan, Peijun Shi, and Autumn Gambles. 2014. Building height estimation from digital surface models (DSMs). International Journal of Remote Sensing.”
to scenes of isolated buildings.

Optical imagery has been employed for height estimation for decades. Both monocular and stereo optical images have been investigated for height applications. In a monocular image, shadows of buildings are used to compute building height (Irvin & McKeown, 1989; Izadi & Saeedi, 2012; Shettigara & Sumerling, 1998). The accuracy of these methods is affected by the quality of detected shadows (especially in distinguishing shadows from water (Shao et al., 2011)) and the building’s context (e.g., adjacent trees and buildings). Before the advent of digital images, stereo pairs of aerial photos have been used for height measurement (Avery & Berlin, 1992). The height of buildings can be measured manually based on stereoscopic parallaxes of roofs and bases of buildings in a stereo model. With the advent of digital stereo imagery and an improvement in computer performance, digital surface models (DSMs) derived from stereo image matching (Birchfield & Tomasi, 1998; Foerstner, 1982; Hirschmuller, 2008) have been widely used for building height extraction. In general, the term DSM represents the Earth’s surface and includes all objects on it (Li et al., 2005), while digital terrain model (DTM) represents the bare ground surface without any objects. The difference between DSM and DTM is described as normalized DSM (nDSM), which represents the height of above-ground objects.

With a given DSM, different strategies have been developed for building height estimation. Most of these methods assume that ground plan maps or previously extracted building footprints are available. Building footprints can be accessed via various sources including cadastral maps, digitized maps from high-resolution images, as well as building boundaries extracted using remote sensing image classification algorithms (Aldred & Wang, 2011), etc. With building footprints and a DSM, the simplest approach to estimate building height is to assume that buildings are built on a certain plane of constant elevation (Lafarge et al., 2008), or assume that terrain elevation is constant in the close neighbourhood of a building (Suveg & Vosselman, 2004). Tack et al. (2012) counted different zonal statistics for given building footprints, such as average or median height. The statistic with the best approximation of the correct building height is selected.
Another major method of building height estimation is to filter the DSM, generate the DTM and calculate nDSM to estimate building height (Casella et al., 2001; Li et al., 2010). Typically, above-ground objects in a DTM usually cannot be removed completely, Mueller et al. (2006) counted only the average of 50% highest values of nDSM over a building rooftop as its height. Finally, there are filters designed specifically for LiDAR point cloud data, which can be transplanted to DSM filtering. LiDAR filter methods include: hierarchic method (Pfeifer et al., 2001), morphological filter (Chen et al., 2007) and additional filters as found in (Sithole & Vosselman, 2004).

In summary, building height estimation based on DSM derived from a stereo image pair is still an unresolved problem and has many challenges. Many methods assume roof-top and/or ground to be flat, which is not true in most cases. The simple morphological filtering method is sensitive to the filter’s window size, while LiDAR filter methods are oriented to point cloud data other than raster cells. Stereo optical imagery, however, is a common and readily available data source for deriving a DSM in urban areas. Therefore, the objectives of this study are: (1) Develop a method to effectively estimate building height from DSMs derived from optical stereo images, with the assistance of building footprints; (2) Based on the estimated height over a building rooftop, analyze building’s structure and generate 3D building models.

4.2 Sources of error in DSM derived from optical stereo images

The accuracy of extracted building height relies on the quality of the DSM been used. Two or more stereo images over the same scene are necessary to calculate the DSM by a process of triangulation, assuming that the imaging parameters and orientation of the sensor are known (Avery & Berlin, 1992). There are factors that affect the accuracy of generated DSM. In urban areas, such factors include occlusions, shadows, little or no texture, and glass walls which are transparent and act as specular reflectors (Gruen, 2012; Zhang & Gruen, 2006). These factors lead to failure during image matching; current commercial software cannot extract elevation values from those failed areas and
therefore, spatial interpolation must be applied to fill the failed areas.

As illustrated in Figure 4-1, satellite sensors retrieve two images of a building from different angles in (a); in (b), the corresponding cross-section of the derived DSM have the elevation changes gradually at the building edges because of the interpolation in occlusions and shadow areas. The “indistinct” building edges, other than discontinuous edges along building boundaries, present the major obstacle for further building height estimation. An example is provided in (c) and (d).

To solve the occlusion problem, Hirschmuller (2008) used the second lowest background value to interpolate and identify occlusion; this operation, while providing more accurate prediction to elevation in most occlusions, cannot change the fact that information is missing in the occluded area. For example, in Figure 4-1(b), this operation leads to the underestimated elevation in the right occlusion. In contrast, multi-ray photogrammetry with more than two view angles (Fraser et al., 2005) provides a solution to the occlusion problem. Multi-ray images captured from different view angles carry highly redundant information of one ground scene, which facilitates the multi-image matching (Zhang & Gruen, 2006) and automatic DSM generation; however, the cost of equipment and the scarcity of the data availability obscure its application. Using commercial software to generate DSM from VHR stereo images (especially one stereo image pair) is currently the main method for DSM production from optical imagery. Consequently, it is still of great importance for a method to estimate building height from DSM derived from a single pair of stereo images.
Figure 4-1. The problem of DSM generation from stereo images. (a) A building is imaged twice by a satellite sensor; (b) the profile of the generated DSM. The solid line is the DSM, the dash lines are the illustrative building and ground. \( Eb1 \) and \( Eb2 \) are the elevation of bare-ground, and \( Er \) represents the elevation of the roof. (c) An example of the DSM extracted from a stereo pair of images using commercial software. The building footprint is overlaid on the DSM. (d) The profile graph of the DSM along the line in (c), from top to bottom.
4.3. Methodology: building-ground elevation difference model (EDM)

4.3.1 The building height estimation model

Building height is the distance from a building’s roof to its base. There are many roof types (e.g., flat, gabled, and hipped) and different ground terrain situations (e.g., flat, slope, and rolling). To simplify the question discussed, a building’s base is defined as the lowest point surrounding the building. As illustrated in Figure 4-1(b), building heights can be defined as:

\[ h(i) = E_r(i) - \min[E_b(j), j = 1,2, \ldots m], \quad i = 1,2, \ldots, n \quad (4-1) \]

where \( h \) denotes building height, \( E_r \) is the elevation of the rooftop, and \( E_b \) is the elevation of the surrounding ground; \( n \) is the number of unique roof elevation values over a building rooftop, while \( m \) is the number of unique ground elevation values in the building neighbourhood. Taking a gabled house built on a flat ground as an example, \( m = 1 \) because the \( E_b \) is constant; \( n >> 1 \) (means far larger than 1) because \( E_r \) changes continuously. For a flat-roofed building built on a slope (see Figure 4-1 (b)), \( n = 1 \) because \( E_r \) is constant and \( m >> 1 \), with \( \min(E_b) \) \( n \) given by its lowest side denoted as \( E_b_2 \) in the figure.

Based on building footprints and DSM, \( E_r \) can be directly obtained, while the estimation of \( E_b \) is difficult due to the “indistinct edge” effect in the DSM (see Figure 4-1 (c)). In the Result Section 4.5.3, it briefly discusses the process to simplify the \( E_r \), calculate \( h \) and reconstruct 3D buildings. This section concentrates on how to estimate \( E_b \) effectively.

4.3.2 Ground elevation: the building-ground elevation difference model (EDM)

Apart from the “indistinct edge” buildings in the DSM, the direct search for \( E_b \) at building’s immediate border is impeded by the building context (e.g., affected by adjacent trees or buildings). To estimate \( E_b \) robustly and effectively, an indirect way is proposed in this study via an elevation difference model.
In order to find the accurate ground elevation $Eb$, this model investigates the elevation change surrounding buildings. Considering that buildings can only affect a certain radius of surrounding grounds in terms of occlusions and shadows, the target of this model is to estimate the radius of effect and then compute the ground elevation. A function is defined to describe the relationship between the distance to building border and the elevation difference $(dE)$ between the building and its neighbour:

$$dE(d) = |\min(0) - \min(d)|, \text{ where } d \text{ is the distance to the building border.}$$

where $d$ is the distance to the building border. $min$ is the minimum elevation in the building’s buffer with given distance $d$, $d=0$ represents the building border; $d$ grows from zero to size $D$.

The plot with $d$ versus $dE$, called an “elevation difference-gram”, is used to analyze elevation change in the building’s neighbouring ground. In a situation where the ground is flat, the $dE$ should be maximized and keep stable after a certain distance. To compute this stable $dE$ and find the “certain distance”, a concept from the semi-variogram model (Curran, 1988; Oliver & Webster, 1986) is adopted. More details about semi-variogram models are given in Appendix F. The spatial dependence in semi-variogram is comparable to the building’s effect on its neighbour’s elevation, while Range in semi-variogram is similar to the “certain distance” describing building’s effect on its neighbouring elevation. As a result, a penta-spherical semi-variogram (Webster & Oliver, 2001) model is selected to fit the elevation difference-gram as following:

$$dE(d) = c_0 \left[ \frac{15}{8} \frac{d}{a_0} - \frac{5}{4} \left( \frac{d}{a_0} \right)^3 + \frac{3}{8} \left( \frac{d}{a_0} \right)^5 \right] \text{ for } d \leq a_0; \text{ and } dE(d) = c_0 \text{ for } d > a_0$$

$a_0$ and $c_0$ are parameters. With a series of $(d, dE)$, $a_0$ and $c_0$ can be computed based on a regression. In this algorithm, $a_0$ is the “certain distance” defined by the radius of building’s effect to its neighbour, called Range; and $c_0$ is the stable $dE$ to describe the building elevation difference on a flat ground, called Sill. An example of the elevation difference model for a building can be found in Figure 4-2(c). With Sill representing the maximum and stable elevation difference, it can be used to calculate the $Eb$ in Equation (4-1) on a flat ground as:
However, more details about the model still need to be clarified before the experiment are conducted. In Equation (4-2), \( d \) increases from 0 to a certain size \( D \) which needs to be big enough for experiments. The size of \( D \) is affected by various factors, such as image resolution, building height, and contextual complexity. In this study, \( D \) is defined as three times the average \( \text{Range} \) of all the buildings after many experiments. Furthermore, to avoid outlier and abrupt elevation (i.e., a hole on the ground), \( \text{min} \) in Equation (4-2) is counted as the elevation where elevation histogram has accumulated a small amount of frequencies \( T\% \) (\( T=5 \) in this study, as based on experimental results).

Given a building boundary, \( \text{min}(d) \) for a series of \( d \) is computed to establish the building’s elevation difference model. A buffer within the distance \( d \) around the building boundary is created, the elevation histogram of the buffer area is established, and the minimum elevation in the histogram is counted as \( \text{min}(d) \). \( dE \) is calculated according to Equation (4-2); a series of \( d \) and \( dE \) are used for regression by Equation (4-3) to compute \( \text{Range} \) and \( \text{Sill} \). The \( \text{Sill} \) is then used to estimate \( Eb \) in Equation (4-4).

### 4.3.3 An example of a building-ground EDM

An example is given to illustrate the building-ground EDM. Figure 4-2(a) shows the aerial image of a building and its surrounding areas; in Figure 4-2 (b), a series of buffers around the building are superimposed on a DSM derived from a stereo pair of aerial images. The gray scale of black to white indicates the elevation range from low to high, respectively. Based on the buffers and the series of \( \text{min}(d) \), the elevation difference-gram is fitted with \( \text{Sill} \) and \( \text{Range} \) calculated. Ground elevation \( Eb \) is calculated according to Equation (4-4).

### 4.3.4 Building-ground EDM adjustment: buildings on slope

The preceding discussion about building-ground EDM and ground elevation estimation is limited to flat ground scenarios. In some cases, buildings are built on slopes, where the ground elevation is not constant. Consequently, the \( \text{Sill} \) value cannot represent the stable elevation difference between building border and ground; in that case, \( \text{Sill} \) is larger than
the actual elevation difference, which underestimates ground elevation $Eb$, see Figure 4-3(a). Therefore, adjustment is needed for the EDM.

Due to the lack of constant ground elevation caused by the presence of a slope, $min(d)$ decreases after $d$ reaches $Range$. However, the speed of decrease directly relates to the slope. Assuming the slope is consistent around the building, the relationship can be defined as:

$$k = \frac{\Delta(dE(d))}{\Delta d}, \text{ where } Range < d < D$$

(4-5)

where $dE$ is the elevation difference, $\Delta$ is the derivative and $k$ is the slope. For a consistent slope, $k$ is constant for the area; for a changing slope, $k$ is used to represent average slope in the area. For $d$ larger than $Range$ and less than $D$, $min(d)$ is used to calculate $k$ via a linear regression. To avoid abrupt elevation change (especially close to $D$), Random Sample Consensus (RANSAC) (Fischler & Bolles, 1981) is adapted to calculate $k$. Under the assumption that slope is consistent, the $Eb$ after adjustment from Equation (4-4) is:

$$Eb = \text{min}(0) - \text{Sill} + k \times \text{Range} \approx \text{min}(0) - \text{min}(d^*) + k \times d^*$$

(4-6)

where $d^*$ is the $d$ where elevation difference ($dE$) is the closest to $\text{Sill}$, among all buffers. Comparing Equation (4-4) and Equation (4-6) can notice that the difference of $Eb$ is the last adjusted item. Furthermore, on a slope, $dE$ is not stable after $Range$; thus, $\text{Sill}$ is not accurate for buildings. For experimental purposes, $\text{Sill}$ is replaced by $\text{min}(d^*)$ and $d^*$ by $\text{Range}$ to improve results accuracy.

In this slope adjustment model, building-ground EDM finds the approximate $\text{Sill}$ first; then $min(d)$ with $d$ larger than $Range$ are used to estimate the slope and adjust the $\text{Sill}$. This adjustment does not affect the buildings built on flat ground where slope is zero. In cases when $k$ is abnormally larger than the average slope ($k > 0.5$ in this study), it is usually an extremely tall building with a larger occlusion; in that case, no adjustment is needed and $Eb$ is estimated via Equation (4-4).
Figure 4-2. The process of building-ground EDM computation with building buffers. (a) Aerial imagery of a building; (b) buffers around the building overlaid on a DSM; and (c) the building-ground EDM of building ground elevation. Note: the scale bar is valid for (a) and (b).
Figure 4-3. Building-ground EDM model on a slope. (a) A building on the slope, with the DSM (solid line) and an illustrative building (dashed line); (b) the building-ground EDM and a fitted line to represent the slope.
Table 4-1. The minimum elevation in buffers and corresponding difference toward building border.

<table>
<thead>
<tr>
<th>$d$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>min(d)</td>
<td>30</td>
<td>24</td>
<td>19</td>
<td>15</td>
<td>12</td>
<td>10.5</td>
<td>19</td>
<td>9.5</td>
<td>9</td>
<td>8.5</td>
<td>8</td>
</tr>
<tr>
<td>$dE$ (d)</td>
<td>0</td>
<td>6</td>
<td>11</td>
<td>15</td>
<td>18</td>
<td>19.5</td>
<td>11</td>
<td>20.5</td>
<td>21</td>
<td>21.5</td>
<td>22</td>
</tr>
</tbody>
</table>

Note: The unit for all the data in the table is meter.
An example of slope adjustment EDM is given in Figure 4-3. A building is built on a slope with $k=0.5$, where building height is 17m and the flat rooftop has an elevation of 30m. The $min(d)$ and $dE(d)$ in buffers are given in Table 4-1. The building-ground EDM is calculated with a $Sill$ of 20.98m; $k$ is 0.5 from the RANSAC regression. $d^* = 8$ has the closest elevation difference (21m) to $Sill$. According to Equation (4-6), the estimated ground $Eb$ is 13m.

4.4 Experiments: building-ground EDM implementation

4.4.1 Study areas and data sources
To implement the building-ground EDM to estimate ground elevation $Eb$, two sites were investigated in London, Ontario, Canada: the University of Western Ontario (UWO) and downtown London (Downtown). UWO is 1.52 km$^2$ with 116 buildings, whereas Downtown is 0.47 km$^2$ with 103 buildings. UWO is built on a hill surrounded by a river (see Figure 4-4(a)) with gentle slopes, while Downtown is a flat area (Figure 4-4 (c)). Average slope for UWO and Downtown is 2.71 degrees and 1.08 degrees, respectively.

The Southwestern Ontario Orthophotography Project (SWOOP) data provides 0.3 m aerial images which were collected in the spring and summer of 2006 (Ontario Ministry of Natural Resources, 2006). These aerial images cover both sites and have 60% along-track overlaps. The aerial images, together with 10 ground control points (GCPs) in each tile have been used to generate a DSM in PCI Geomatica. The derived DSM has 0.6 m resolution. A building footprint vector map was produced by City of London, Ontario for 2006. Due to recent construction, a few buildings mismatch between the vector map and aerial images; they are removed from this study. In order to collect ground truth testing points, 49 roof-top height sample points were measured in UWO in March, 2011 and 46 sample points were surveyed in Downtown in June, 2012 (see points labeled as crosses in Figure 4-4 (b) and (d)). These samples cover buildings with different heights, from low-rise buildings to extremely tall buildings. The laser rangefinder, with accuracy of 0.3 m on distance, was used to measure heights for sample buildings. For a given point, the distance from a roof edge point to the vertical ground was measured and recorded as the
height for the building at that point.

The DSM of the Downtown study site includes crucial elevation errors around extremely tall buildings, due to large image parallax and long shadows. To evaluate the building-ground EDM and avoid the large DSM error as input, LiDAR data was used to subset the study area. A mask was created for areas where elevation difference between LiDAR and DSM were larger than 3m. The LiDAR point cloud with 0.8-0.9 point/m² was re-sampled to the same spatial resolution (0.6m) as the DSM for masking purposes. The study area, as refined through the use of masks, includes 46 surveyed rooftop points.

4.4.2 Experiment steps

As shown in Figure 4-5, the flowchart of height estimation includes several steps. The key step is the building-ground EDM to calculate $E_b$, as introduced in the Methodology Section 4.3.1. The two input data layers include the building footprints and the DSM derived from aerial images. Building footprints and their buffers were used to search for elevation over a building rooftop ($E_r$ in Equation (4-1)) and minimum elevation in each buffer, respectively. The minimum elevations are input into the ground elevation building-ground EDM model to estimate $E_b$. As a result, the building height can be estimated according to Equation (4-4) or Equation (4-6). After validation, the building height histogram is analyzed to reconstruct 3D buildings.
Figure 4.4. The two study sites and their DSM. (a) Aerial image of UWO and building footprints; (b) DSM of UWO and sample building height points; (c) and (d) are aerial image and DSM for Downtown.
Figure 4-5. Ground elevation estimation and building height extraction workflow.
4.5. Result: Roof-top point height evaluation and 3D building reconstruction

4.5.1. Performance evaluation: building height error analysis

To evaluate the performance of the building-ground EDM model in this study, building height estimated from the original EDM (assuming the ground is flat near a building) and the slope adjustment EDM are both implemented. The result is also compared with the traditional approach, by subtracting a DTM from a DSM. Two DTM sources are investigated: one is from an independent organization and another is obtained by filtering the current DSM. The independent DTM comes from Provincial Digital Elevation Models v2.0.0 (OGDE), which has a 10m horizontal resolution and 5m vertical reliability (Ontario Ministry of Natural Resources, 2004). This DTM was re-sampled to the same resolution as DSM. The second DTM data are generated by filtering DSM. This filtering process is conducted in a commercial software LiDAR Analyst (A Visual Learning Systems, 2005). The “Bare Earth Extraction” filters a DSM into DTM, using the “Hierarchical Spline Interpolator”. To refine the second DTM, elevations over building rooftops are replaced with their minimum elevation after filtering.

Surveyed rooftop height points are used to validate methods. To estimate height at sample points according to Equation (4-1), $E_r$ is the estimated elevation from the DSM while $E_b$ is estimated from four approaches: (a) DSM filtering, (b) independent DTM, (c) the proposed original EDM, and (d) the slope-adjusted EDM. Errors are counted as the estimated height from different approaches, subtracting ground surveyed height. Mean error (ME), mean absolute error (MAE), standard deviation (STD), and root mean squared error (RMSE) are reported for different methods in Table 4-2.

For the four approaches implemented in Table 4-2, the building height from the “DSM Filter” method by LiDAR Analyst has an error about 2m for both sites. The MAE of height error for UWO is smaller than that in Downtown. One explanation is that building plan area density is greater in Downtown; even with the hierarchical strategies, it is
difficult to distinguish large, connected buildings from the ground using the filtering process. Compared with the “DSM Filter” in Downtown, the error for the “Independent DTM” method drops to nearly a half. It demonstrates that an independent DTM can add useful information to suppress the height errors. With the proposed EDM methods, building height is estimated with comparable accuracy to “Independent DTM” method, but without using an extra DTM.

Comparing the “EDM” methods before and after slope adjustment, the error has reduced after adjustment. The MAE decreases about 0.3m in UWO but only 0.1m for Downtown. The slope adjustment is more effective for the UWO site, as buildings were built on the existing sloping landscape, in contrast with the flat terrain of Downtown where no slope adjustment is needed. Before adjustment, both sites have positive residual in ME. Original EDM sometimes overestimates the building height because of its underestimation of ground elevation. After adjustment, the ME value is still positive in slope study site, but becomes negative in flat study site; the ME of both sites decreases after slope adjustment. Overall, the proposed “EDM” methods estimate building height with equal or better accuracy to the “Independent DTM” method, even though the former does not use an extra DTM as input.
<table>
<thead>
<tr>
<th>Method</th>
<th>DSM Filter (LiDAR Analyst)</th>
<th>Independent DTM</th>
<th>EDM (original)</th>
<th>EDM (slope adjusted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation</td>
<td>DSM -DTM</td>
<td>DSM -DTM</td>
<td>DSM -Eb</td>
<td>DSM -Eb</td>
</tr>
<tr>
<td>DTM</td>
<td>Filtering DSM</td>
<td>Independent source</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Study sites</td>
<td>UWO Downtown</td>
<td>UWO Downtown</td>
<td>UWO Downtown</td>
<td>UWO Downtown</td>
</tr>
<tr>
<td>ME (m)</td>
<td>1.73</td>
<td>1.95</td>
<td>-0.69</td>
<td>-0.15</td>
</tr>
<tr>
<td>MAE (m)</td>
<td>1.97</td>
<td>2.25</td>
<td>1.56</td>
<td>1.45</td>
</tr>
<tr>
<td>STD</td>
<td>1.96</td>
<td>2.61</td>
<td>1.27</td>
<td>1.35</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.94</td>
<td>2.61</td>
<td>1.26</td>
<td>1.33</td>
</tr>
</tbody>
</table>
4.5.2 Building-ground EDM elevation model in dense urban areas

Dense urban areas with more buildings in the complex environment make the height estimation more challenging. But the EDM method performs consistently for both study sites; the error is even smaller for Downtown than UWO. This is because EDM only searches for the minimum elevation around the buildings; as soon as the buffers cover open areas such as roads or parking lots, the ground elevation can be correctly estimated. As shown in Figure 4-6, buildings connect to each other and the main roads are occluded by shadows (see 6(a)); most parts of the roads are filled with inaccurate elevations (see 6(b)). However, \( \min(d) \) in building buffers, will reach a stable value if any side of the building is adjacent to a backyard or a parking lot.

In contrast, the traditional morphological filters still fail to remove the inaccurate elevations in neighbouring areas to buildings, as it is difficult to select an appropriate window size during filtering. This problem is solved using the EDM method, since the building-ground EDM will self-adaptively determine the distance when the ground is touched. This distance is the \( \text{Range} \) for a building.
Figure 4-6. An example of building height estimation at Downtown. (a) A subarea of the aerial image in Downtown; and (b), the corresponding DSM, buildings and buffers.
4.5.3 The 3D building reconstruction for typical buildings

If the $Eb$ for a building can be effectively estimated using the EDM, a 3D building model can be generated. According to Equation (4-1), the height of a building can vary from one constant value to many values. To simplify the question, three typical cases of building are discussed. Englert and Gulch (1996) had categorized the building into different types. Considering that reconstruction is beyond the scope of this study, three typical roof types are adopted for consideration: the flat-roofed building, the gabled house, and the complex buildings with several flat roofs.

The building height histograms provide a tool for a semi-automatic 3D building reconstruction. As various types of buildings possess different histogram prototypes (see Figure 4-7(b)), the histogram was utilized to recognize the building types. Flat buildings are typically represented by a clearly unimodal histogram. For multi-roof buildings, each peak represents a group of heights in the building footprint. Therefore, valleys in histogram can be used as thresholds to separate buildings into multiple parts.

In an actual case of gabled houses (see Figure 4-7 (d)), the height histogram usually turns out to be a bimodal. In a DSM derived from stereo images, a gabled roof usually lacks enough details to reflect continuous change of roof elevations, due to limitations on image resolution and matching techniques. As a result, elevations over a gabled house roof move into two groups: the eaves and ridges. However, this bimodal shape is quite different from that of the multi-roof type, as looking into the histogram valley. The local minimum for a gabled house histogram is not as low as the multi-roof type because considerable proportions of elevation exist in between the two peaks (eaves and fastigium). With the given type and height, the 3D building can be reconstructed, as illustrated in Figure 4-7 (e).
Figure 4-7. The 3D reconstruction of three types of buildings. (a) Parametric models of different building types. (b) The histograms of parametric building models. (c) Example buildings with the derived DSM superimposed on the aerial images. (d) The histograms
of the example buildings shown in (c). (e) The 3D models of the three example buildings. Column (i) is flat-roof building, (ii) is gabled house, and (iii) is building of combinations.

4.6. Conclusion

In this chapter, a building-ground elevation different model (EDM) is proposed and building height is estimated. In urban areas, factors affecting generation of DSM from stereo imagery include occlusions, shadows, little or no texture, and transparent glass walls that act as specular reflectors, etc. The DSM derived from such high resolution optical stereo images by current commercial software resulted in inaccurate surface elevations. The surface elevations “overflow” from a building to its neighbours which blurs the building edges in the DSM, submerges building immediate neighbours and blocks direct access to ground elevation. To solve this problem and to estimate building height, an indirect method called building-ground EDM model has been proposed to first estimate the ground elevation and then to calculate building height.

The EDM transplants the concept of spatial dependence and the influence range from semi-variogram. The elevations of a building’s immediate neighbours are affected by the building, while a certain distance is needed to get rid of this effect. This certain distance is the Range, where stable ground elevation usually is reached. To search the Range using the building-ground EDM model, building buffers are employed to trace the building-ground elevation difference change. Then Range is fitted from building-ground EDM via a penta-spherical semi-variogram model. For buildings on slopes, average slope is estimated and ground elevation is adjusted according to the slope.

Surveyed rooftop height points from two study sites, one on gentle slopes and the other on flat terrain, are used to evaluate the performance of the EDM. Other traditional approaches estimating building height by subtracting DTM from the DSM are compared with the proposed EDM.

Several conclusions are drawn from the results: (a) From the comparison of EDM and other height extraction methods, the results show that the slope adjusted EDM were more
accurate than the DSM filter approach. Even without an extra DTM, the slope-adjusted EDM method produced comparable or better accuracy to the independent DTM method. The height error of the testing points is less than 1.5m. (b) Comparing the result of original EDM and slope adjusted EDM, slope adjustment can remove the underestimation of bare-ground elevation; thus reducing the overestimation of building height. The slope adjustment effect is more effective in steeper study area. (c) The EDM reaches similar accuracy in both open, spacious areas and high-density urban areas. That is because the EDM can self-adaptively find the distance where the effect from neighboring buildings can be removed and ground can be found. (d) After the ground elevation is estimated, typical 3D buildings are reconstructed in a semi-automatic manner, based on the relationship between building types and their histogram patterns. Three typical building types are discussed and examples are given.

During the experiments and analysis, some aspects need further research and exploration. First of all, because the DSM is the major input to the EDM, the accuracy of DSM (especially over building rooftops) has a large impact on the accuracy of the extracted building heights. Extremely tall buildings usually fail to be accurately matched in image matching because of their large image parallax and long shadows. The quality of DSMs derived from single stereo imagery pairs by current commercial software still has room for improvement. Furthermore, only three types of buildings are investigated in 3D reconstruction. As three building types admittedly cannot represent all buildings in a complex urban context, reconstruction covering more building types would be well-suited to further research. Finally, topics relating to the current research deserve further exploration, such as the relationship between Range and image resolution, view angles, etc, and the building footprints registration on DSM.

References


Ontario Ministry of Natural Resources. (2004). *Ontario Ministry of Natural Resources Digital Elevation Model (DEMS) - 10 and 20 metre coverage*. Peterborough, ON: Ontario Ministry of Natural Resources [producer(s)], Ontario Geospatial Data Exchange (OGDE) [distributor(s)].


Chapter 5: A Multi-criteria Evaluation Method for 3D Building Reconstruction*

5.1 Introduction
A three-dimensional (3D) building model is not only an important product itself, but also a valuable input for other applications. 3D buildings are the major objects employed for virtual environment visualization, natural disaster risk management (Geiß & Taubenböck, 2012), city planning, and location-based spatial analysis. The accuracy with which a 3D building is generated determines its scope of application. Without a credible accuracy report, building models will be either treated as correct or considered as useless because the quality of digital reconstruction is unknown (Elberink & Vosselman, 2011).

Recently, 3D buildings are reconstructed with high accuracy and wall details using advanced methods and various data sources (Haala & Kada, 2010; Henn et al., 2013; Huang et al., 2013). To evaluate the accuracy of a reconstructed building, straightforward metrics are developed from 2D evaluation. Most metrics only evaluate the rooftops, ignoring the walls. Simple metrics such as detection and quality percentage (McKeown et al., 2000) can only be used to partially evaluate buildings. A more complicated evaluation system is expected to enhance the evaluation process and subsequent comparison between various 3D building reconstruction methods.

In Chapter 2, a multi-criterion evaluation system was proposed to assess the extracted 2D building footprints, using three components (matched rate, shape similarity, and positional accuracy) to evaluate 2D building footprints. In this Chapter, 3D building evaluation methods based on the previous Chapter 2 is explored. Considering that 2D and 3D buildings are essentially different in form, new evaluation metrics are required. As a result, the objectives of this research are:

(1) To review current 3D building evaluation under a new framework.

(2) To develop a multi-criteria evaluation system for reconstructed 3D buildings.

(3) To implement a simple 3D building reconstruction to test the proposed evaluation system.

5.1.1 The Status of 3D Building Reconstruction

A range of views can be used to categorize 3D building reconstruction from remotely sensed data. From a methodological perspective, data-driven and model-driven methods are widely used. In data-driven methods, point clouds and digital surface models (DSMs) are segmented and grouped, features are recognized, and 3D models are built accordingly (Lafarge & Mallet, 2012; Zhang et al., 2012). In model-driven methods, typical roof types are predefined while input point clouds or DSMs are fitted to predefined roof types for modelling (Henn et al., 2013; Huang et al., 2013). In terms of data sources, LiDAR point cloud and multiple-view stereo imagery are two major data sources for 3D reconstruction. LiDAR points with elevation values are directly used for DSM generation or planar segmentation (Henn et al., 2013; Huang et al., 2013). Alternatively, stereo imagery can be matched according to multi-image matching algorithms (Hirschmuller, 2008; Zhang & Gruen, 2006); then, 3D buildings can be reconstructed (Schmid et al., 2012).

Furthermore, there is a trend to reconstruct buildings with increasing level of details (LoD) (Gröger & Plümer, 2012) in order to describe buildings more accurately. Previous reconstruction methods treated roofs as flat; subsequent methods were improved to consider popular roof forms (i.e., gable, hip, etc.). Later, more advanced methods account for roof details such as chimneys and air-conditioner. Current approaches focus on the inclusion of wall facades such as windows and doors in the course of reconstruction (Haala & Kada, 2010). The LoD improvement is directly related to improvements in sensor ability with higher spatial resolution and more view angles. Aerial imagery and LiDAR data have been used to reconstruct building roofs; however, such data lacks the ability to reconstruct wall facades, even when oblique view images are used (Petrie, 2009). Recent terrestrial data, such as terrestrial LiDAR data (Pu & Vosselman, 2009), handheld digital cameras (Bhatla et al., 2012), or video streams (Brilakis et al., 2011;
Pollefeys et al., 2008), can act as complementary data and provide support for a new trend towards wall façade reconstruction. For a detailed review of 3D building reconstruction, readers may refer to Haala and Kada (2010).

5.1.2 The Framework to Review 3D Building Evaluation

Although many efforts have been made in 3D building reconstruction from remotely sensed data, the evaluation process has not been studied thoroughly. Similar to other object-based evaluation, 3D building evaluation involves assessing a sample object from 3D building reconstruction and its corresponding reference object. The accuracy issue can then be converted into comparison between the sample object and reference object: the less difference between the sample object and the reference object, the higher the accuracy of the reconstruction process. To quantify this comparison, both sample and reference objects are primarily described by one or more features, with subsequent metrics derived based on these features as in Equation (5-1),

\[ m = f(V(r), V(s)) \]  

(5-1)

where \( m \) is the evaluation metric, \( f \) is the comparison function, \( V \) is the vector to describe a feature, \( s \) and \( r \) represent sample object and reference object respectively. For example, in a simple metric “centroid distance” that defines distance between centroids of 3D sample building and reference building, \( V \) is the 3D vector of building centroid and \( f \) is the Euclidean distance between the two 3D points. For given sample data \((s)\), the evaluation process is to generate proper reference data \((r)\), define evaluation feature \((V)\), and identify comparison function \((f)\), in order to accurately and effectively describe the relationship between the 3D sample reconstruction building and the reference building.

Reference data \((r)\). The direct buildings reference data comes from image digitalization with higher level of details in 3D virtual environment. Photogrammetry and stereo model is a popular method to digitalize accurate buildings in a 3D environment. Although expensive in terms of time and cost, they are the primary reference data. In addition, LiDAR point clouds or digital aerial imagery are also used as indirect reference data for evaluation with lower cost. LiDAR points with elevation information can validate reconstructed roof-tops (Akca et al., 2010; Elberink & Vosselman, 2011), while aerial imagery can collect positive or negative evidence of roof facet consistency (Boudet et al.,
Finally, 2D databases (i.e., cadastral maps) also provide reference data for evaluation conducted on a 2D plane (Huang et al., 2013) or simple prismatic buildings.

**Evaluation features (V).** It is critical to define an effective evaluation feature that can distinguish reference building from sample building. *Area* and *Volume* are the intuitive features for building from human vision; therefore, the most commonly used features for building evaluation are Area in 2D space and Volume in 3D space. True positive (TP), false positive (FP), true negative (TN), and false negative (FN) (McKeown et al. (2000) are used to develop metrics (Landes et al., 2012) based on 2D pixels or 3D voxels (Karantzalos & Paragios, 2010). In addition, locational features are another important category; popular locational features include centroids, corners, and mid-points of edges (Akca et al., 2010). Finally, evaluation features can be defined in either 2D or 3D space. A 2D feature is based on projection on a certain plane such as axis-planes (Huang et al., 2013); the 3D features use the true building shape to derive features such as point position and normal vector (Landes et al., 2012).

**Comparison functions (f).** The difference and similarity comparison functions are two opposite function groups. Difference functions describe the discrepancy between reference data and sample data, and similarity functions define the degree of resemblance. In (McKeown et al., 2000), 3D buildings are evaluated by various metrics, where detection percentage and quality percentage are under the similarity function category, while the omission error and commission error are difference functions. Comparison functions can be normalized with metric value between 0 and 1. Most ratio functions, including quality percentage or branching factor (McKeown et al., 2000), are normalized; distance functions such as shift, rotation and scale on X, Y, and Z directions (Akca et al., 2010; Ameri, 2000) are usually not normalized. Comparison functions can be computed independently or interactively, depending whether features from reference building and sample building interact with each other.

Progress in the 3D building evaluation has been made by current studies. McKeown et al. (2000) investigated the accuracy of 3D features and provided the basic definition of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) inherited from the traditional classification accuracy evaluation. In Boudet et al. (2006), a facet
quality self-diagnosis analysis is proposed to separate buildings into different groups and evaluate them via supervised classification. In a survey and research conducted by Sargent et al. (2007), factors that are important in evaluating a building’s 3D model accuracy are categorized, including geometric fidelity, relative positional accuracy, and absolute positional accuracy, together with their sub-categories. In Akca et al. (2010), the least squares 3D surface matching method is utilized to evaluate the distances (translational, rotational, and scale) between the reconstructed rooftop and LiDAR points. In Landes et al. (2012), the quantitative method uses completeness, normal to the planes, and offsets between the reference points and sample points to evaluate the quality and accuracy of planar clusters in 3D models. Landes et al. (2012) employ an error map describing the surface mismatch distance of each plane, which is also introduced to visually assess reconstruction accuracy. In Mohamed et al. (2013), an approach that assesses buildings in 1D, 2D and 3D is proposed, and traditional quality indices such as quality and completeness are derived from building planes and volumes.

There are other evaluation strategies that concentrate on the input data quality (Baltsavias, 1999), errors related to both the instruments and the object under study, environmental errors and methodological errors (Reshetuk, 2009), or the error propagation in 3D building reconstruction procedure (Elberink & Vosselman, 2011; Vosselman, 2012). While the importance of these evaluative strategies for explanation of error occurrence is acknowledged, the specific details of these evaluations are beyond the scope of this study and will not be discussed within this thesis.

In summary, 3D building reconstruction evaluation is still a challenge as 3D reconstruction is not as well-developed as 2D extraction methods. Evaluation metrics for 3D reconstruction are mainly derived from 2D metrics such as completeness, detection rate, quality (Landes et al., 2012; McKeown et al., 2000; Mohamed et al., 2013), and positional distance metrics (Akca et al., 2010; Ameri, 2000). The difference for 2D buildings and 3D buildings is not well-presented, considering that 2D buildings exist as a footprint in a 2D plane, while 3D buildings exist in a 3D space with complicated surface and structure. Moreover, current evaluation methods merely assess the rooftop (Akca et al., 2010; Elberink & Vosselman, 2011; Vosselman, 2012) rather than entire 3D buildings.
A rooftop can only represent an entire building if operating under the assumption that a building has no wall detail. This assumption is not true for buildings with overhang-roofs or complicated shapes.

This difference between 2D and 3D building evaluation can be partially solved by evaluating roofs and facades separately (Landes et al., 2012). However, a 3D building is different from its 2D projections because the 3D surface information is lost during the projection; therefore, a direct comparison in 3D space is necessary. Additionally, matching all 2D planes (Mohamed et al., 2013) “facet-to-facet” encounters the automatic registration problem. Due to different data structures used to organize the 3D shape and inconsistent building parts, it is difficult, or sometimes impossible, to find homologous planes between the sample and the reference buildings. With the development of terrestrial sensors, including terrestrial laser scanning, ground-based multi-angle cameras (Bhatla et al., 2012), and video-grammetry (Brilakis et al., 2011), more accurate building models can be reconstructed with higher accuracy and details on both the roofs and the walls. The increasing impact of wall details on reconstruction accuracy challenge current 3D building evaluation methods, requiring the development of an advanced evaluation system.

5.2 Methodology
To build a multi-criteria evaluation method, Boudet et al. (2006) considered three types of error in 3D building models: the non-existence of the corresponding building, inaccuracy in shape description, and the geometrical inaccuracy of a 3D facet. These three aspects of error are considered in the proposed evaluation system that covers three components. The first component describes volume accuracy: it measures the percentage of volume that is correctly or mistakenly reconstructed. It is based on TP, TN, FP, and FN (Karantzalos & Paragios, 2010; McKeown et al., 2000), but uses voxels rather than pixels. The second component is surface accuracy, which measures the similarity between surfaces of the 3D sample building and reference building. The third component defines point accuracy, which calculates the distance between corresponding feature points of the sample and the reference data. Commonly used feature points include...
Reconstructed 3D buildings are evaluated using three components from different perspectives. *Volume* accuracy describes the size similarity between buildings. Comparison functions (i.e., correctness, completeness, and Quality) are computed based on evaluation features including the percentage of overlapped (TP), missed (FN), and over-reconstructed (FP) volumes. For buildings in vector format, 3D intersection (Mohamed et al., 2013) is used to calculate TP, FN, and FP. For buildings in voxel format, the “voxel-in-volume” test is used for all voxels. The intersection of 3D vectors is not feasible for complicated buildings with thousands of edges; in contrast, a “voxel-in-volume” test for each voxel is stable but inefficient. A random sample method is a compromise between 3D vector intersection and the “voxel-in-volume” test. A random sample is labelled as TP, FN, or FP, depending on whether it is inside or outside the reference building and the sample building. As illustrated in Figure 5-1 (a), a random point is considered to be inside a building only if its six rays of orthogonal directions intersect with the building walls at the same time.
Table 5-1. Summary of the three components for accuracy evaluation

<table>
<thead>
<tr>
<th>Components</th>
<th>Summary</th>
<th>Comparison functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>The volume difference between corresponding 3D buildings</td>
<td>completeness, correctness, quality percentage, etc.</td>
</tr>
<tr>
<td>accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface</td>
<td>The surface similarity between corresponding 3D buildings</td>
<td>SPHARM coefficients RMSD* (Brechbühler et al., 1995), spin image (Johnson &amp; Hebert, 1999) after sphere parameterization, etc.</td>
</tr>
<tr>
<td>accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Point accuracy</td>
<td>The positional accuracy for feature points between a sample building and its reference building.</td>
<td>RMSD, Mean, and standard deviation for Euclidean distance</td>
</tr>
</tbody>
</table>

*RMSD is “root mean square deviation”.*
The point accuracy describes the absolute positional offset from the reference building to the sample building. Point accuracy is important for applications with particular concern regarding building locational accuracy. Point accuracy is computed using Euclidean distances between 3D points of sample and reference buildings. Comparison functions for points include distance mean, standard deviation, and root mean square distance (RMSD). To register feature points from a reference building to a sample building, point contextual information is used to remove incorrect candidates and enhance registration. For example, corner points are registered dependent on not only their coordinates, but also the norms of adjacent facets, as illustrated in Figure 5-1 (b).

Surface accuracy measures shape similarity by comparing two building surfaces in a 3D space. Currently there is no true 3D surface comparison metric for buildings. Most studies use point clouds to represent 3D surfaces, such as the surface matching algorithm (Akca et al., 2010; Gruen & Akca, 2005) and iterative closest point methods (Besl & McKay, 1992; Zhang, 1994). The projection of a 3D surface onto several 2D planes is another strategy; however, surface details are lost during the projection. In addition, it is difficult to register homologous planes automatically between two 3D surfaces for “facet-to-facet” comparing methods (Mohamed et al., 2013). As a result, this chapter proposes a direct 3D building surface comparison method.
Figure 5-1. Illustration of 3D building evaluation. (a) **volume** accuracy: point-in-volume detection with six rays of orthogonal directions; (b) **point** accuracy: a corner point is registered based on its coordinates and adjacent facet norms; (c) **surface** accuracy: two buildings of different shapes are mapped onto a unit sphere by sphere parameterization.
In “facet-to-facet” surface comparison, it is difficult to search for corresponding facets due to data structure difference and building part mismatch between sample buildings and reference buildings. To avoid searching for corresponding facets, a global comparison directly compares both buildings in their entirety, rather than individual facets. To make two building surfaces comparable, both are mapped onto a unit sphere as long as they are topologically equivalent to a sphere (Brechbühler et al., 1995), as illustrated in Figure 5-1 (c). Buildings with holes (e.g., an inner courtyard) need to be manually separated into parts; each part is then compared accordingly. The process of mapping from a closed surface to a unit sphere, called “sphere parameterization”, is a constrained optimization problem with many implementations, such as “temperature” diffusion (Brechbühler et al., 1995; Chung et al., 2007) and the control of area and length distortions (Shen & Makedon, 2006). After two building surfaces are mapped onto the same sphere, they can be compared on the unit sphere with existing metrics such as surface curvature index (Hebert et al., 1995), spin image (Johnson & Hebert, 1999), and spherical harmonic (SPHARM) (Brechbühler et al., 1995; Chung et al., 2007). More details about Spherical harmonic (SPHARM) representation for 3D objects are given in Appendix G.

SPHARM is invariant to translation, scale and rotation (Brechbühler et al., 1995); thus, it is selected in the proposed system to compare 3D building surfaces. SPHARM decomposes a 3D shape into a complete set of basic functions (Brechbühler et al., 1995) (spherical harmonics). Using the coefficients of all spherical harmonics, the 3D object can be reconstructed. The comparison of two 3D buildings can be achieved using their SPHARM coefficients after normalization (Gerig et al., 2001):

\[
\text{RMSD} = \sqrt{\frac{1}{4\pi} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \|c_{1,l}^m - c_{2,l}^m\|^2}
\]

(5-2)

where RMSD is the root mean squared distance between SPHARM coefficients of the two buildings, \(c\) is the normalized spherical harmonics coefficients, \(l\) is degree, \(m\) is the order. The larger the degree \(l\), the closer \(c_l\) can represent the 3D building. To better
describe discontinuous features (building edges), a weighted linear combination of spherical harmonics (Chung et al., 2007) is adapted.

5.3 Experiments and evaluation process

5.3.1 Three dimensional building reconstruction from LiDAR data

To test the proposed multi-criteria evaluation method, 3D buildings are initially reconstructed from airborne LiDAR data. The study area is the campus of the University of Western Ontario, Ontario, Canada. The aerial LiDAR data captured in 2006 has point density of 0.8-0.9 points per square meter. Buildings are reconstructed in 3D via a few steps, as illustrated in Table 5-2. LiDAR point clouds are initially pre-processed and interpolated to raster images. Edges in the raster images are detected with Canny operator; from there, edges are then vectorized, closed up, and filtered by a 2D building footprint map. Flat and pitched roof structures are identified according to roof slope. The resultant 3D building models are further reconstructed separately for flat and pitched roofs, where edges on pitched roofs re-clustered based on aspect (Zhao, 2013). A triangulated irregular network (TIN) is generated for pitched roofs.

5.3.2 Three dimensional building evaluation via the multi-criteria evaluation system

After the 3D buildings of interest are reconstructed from the LiDAR data, they are evaluated through comparison to their respective reference buildings. Reference buildings are manually edited, based on LiDAR and 30cm high-resolution aerial photos. 3D building edges are digitalized in AutoCAD Map3D, which supports 3D feature editing over geo-referenced data. The horizontal edge position is mainly determined by aerial photos and assisted by LiDAR data; edge elevation comes from LiDAR data. The level of detail used for digitalization is depended on data spatial resolution. Specifically, building roof details larger than 2m by 2m in size (about twice of the LiDAR point gap) are digitalized in the reference buildings.

The three accuracy components are implemented as shown in Figure 5-3. The Matlab code implemented the three component comparison is given in Appendix H. For the volume component, TP, FP, TN and FN are computed based on the sample building and
the reference buildings. As illustrated in Figure 5-4, two thousand random sample points are labelled as different categories (TP, FP, TN, FN), depending on their “point-in-volume” relation with the sample building and reference building. For instance, a random point inside both sample and reference building is labelled as TP. According to (McKeown et al., 2000), 3D evaluation metrics including completeness, correctness, and quality are calculated based on the number of points in each category.

For the surface component, the building surface is initially mapped onto a unit sphere using the diffusion of “temperature” algorithm (Chung et al., 2007), where the building surface is treated as a “source” and a sphere containing the building is treated as a “sink”. With the heat sink and source, isotropic heat diffusion is performed. After a sufficient amount of time, a steady state is reached, which is equivalent to solving the Laplace equation. The path for each pixel from the inner “source” (building surface) to the outer “sink” (the unit sphere) is determined. Spherical harmonics coefficients are then calculated and the 3D building is deformed from the unit sphere, as illustrated in Figure 5-5. Using the spherical harmonics coefficients for sample and reference buildings, the distance to describe the shape similarity between reference and sample buildings is computed according to Eq. (2).

For the point component, corner points are selected to evaluate the 3D building positional accuracy. The corner points from each sample building and reference building pair are automatically registered. The registration is dependent on distances between corner points and norm vector angles of corresponding adjacent facets. Differential vectors are then calculated based on point pairs, as illustrated in Figure 5-6.
Figure 5-2. The procedure of 3D building reconstruction from LiDAR data. (a) The rasterized LiDAR data is overlaid by building footprints; (b) edge detection and post-processing; (c) classification of roof segments draped on slope image (d) reconstructed 3D buildings.
Figure 5-3. Framework for 3D building evaluation. Note that each component has its evaluation features and comparison functions.
Figure 5-4. Random sample points are used to evaluate *volume* accuracy. (a) The reference building and random points. (b) The sample building and random points. Blue points are outside building, while red points are inside building. Note: the boundary of the box is determined by the minimum and maximum coordinates for both 3D buildings.
Figure 5-5. 3D buildings sphere parameterization and SPHARM. (a-d) is the process of “temperature diffusion” from the reference building model in (a) with temperature (t) =1, to the final unit sphere surface in (d) with temperature as t=-1. (e) is the building model generated via SPHARM coefficients at degree (l) 80. (f-j) is the same process as (a-e) but for the corresponding sample building. For details about the temperature (t), please refer to Appendix E.
Figure 5-6. The *point* accuracy evaluation for an example building. A difference vector is from a point on the reference building to its corresponding point on the sample building. Note: some points do not have matching point in the reference building due to wrong or different data structure. The unit of measurement is meter.
5.4 Results and Discussion

After reconstruction and evaluation of the buildings, example buildings in the study area with their evaluation results are reported in Table 5-2. Completeness, correctness, and quality are computed to describe 3D building volume accuracy. The range for these three metrics is 0 to 1; the closer the metric value is to 1, the more accurate the reconstruction result. The surface accuracy is implemented as the RMSD of SPHARM coefficient distance. This RMSD is not normalized; the lower value indicates less outline discrepancy and higher reconstruction accuracy. To make this RMSD comparable to the other two components, it is normalized as:

\[
rate = 1 - \frac{RMSD}{MAX}
\]  

(5-3)

where rate is the normalized shape similarity, MAX is the RMSD between a unit cubic and a unit sphere (which is 7.9 in this experiment). The point accuracy, reported as the RMSD for the vector (dx, dy, dz), is the corner difference between the sample building and the reference building; a lower differential vector means less location shift between sample building and reference building corners.

Although it is an important metric, volume accuracy values are close to its theoretical upper bound (e.g., building “UCC” in Table 5-2 has Correctness 0.95 while the upper bound is 1). The current methods are outstandingly accurate to the point that there is limited room for comparison to future emerging methodologies. Furthermore, standard deviation (STD) for volume metrics is lower than the other two groups of metrics. The lower STD suggests that volume accuracy metrics perform relatively evenly on all buildings, with less contrast evident between the accuracies of individual buildings.
Table 5-2. The evaluation result for example buildings with multi-criteria.

<table>
<thead>
<tr>
<th>Building Name</th>
<th>Volume</th>
<th>Surface</th>
<th>Point</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Completeness</td>
<td>Correctness</td>
<td>Quality</td>
</tr>
<tr>
<td>Alumni Hall</td>
<td>0.89</td>
<td>0.86</td>
<td>0.77</td>
</tr>
<tr>
<td>Delaware Hall</td>
<td>0.94</td>
<td>0.88</td>
<td>0.83</td>
</tr>
<tr>
<td>Health Sci. Bldg.</td>
<td>0.85</td>
<td>0.89</td>
<td>0.77</td>
</tr>
<tr>
<td>Ivey School</td>
<td>0.93</td>
<td>0.94</td>
<td>0.88</td>
</tr>
<tr>
<td>Law School</td>
<td>0.91</td>
<td>0.89</td>
<td>0.82</td>
</tr>
<tr>
<td>Middlesex College</td>
<td>0.89</td>
<td>0.89</td>
<td>0.80</td>
</tr>
<tr>
<td>Social Sci. Centre</td>
<td>0.95</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td>Somerville House</td>
<td>0.83</td>
<td>0.82</td>
<td>0.71</td>
</tr>
<tr>
<td>Talbot College</td>
<td>0.95</td>
<td>0.93</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>UCC</strong></td>
<td>0.91</td>
<td>0.95</td>
<td>0.87</td>
</tr>
<tr>
<td>University College</td>
<td>0.83</td>
<td>0.83</td>
<td>0.71</td>
</tr>
<tr>
<td>Weldon Library</td>
<td>0.95</td>
<td>0.91</td>
<td>0.87</td>
</tr>
<tr>
<td><strong>STD</strong></td>
<td>0.04</td>
<td>0.04</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Note: \(^a\) the building UCC is used for illustration in Fig 4-6. \(^b\) Rate is the normalized RMSD value, calibrated with the maximum RMSD difference from a unit sphere and a unit cubic. \(^c\) STD is the standard deviation for each metric.
Figure 5-7. Examples of 3D building shape comparison. The left column is the reference 3D buildings, the right column is the extracted 3D buildings. (a) and (b) are Alumni Hall as listed in Table 5-2. (c) and (d) are Middlesex College, (e) and (f) are Law School.
Figure 5-7 (Continue). Examples of 3D building shape comparison. The left column is the reference 3D building, the right column is the extract 3D building. (g) and (h) are Ivey School as listed in Table 5-2.
The *surface* accuracy, described by SPHARM RSMD and its normalization, introduces a new dimension of 3D building evaluation. It is unique in terms of *volume* and *point* accuracies because it is invariant to building translation, scale and rotation; thus, the SPHARM RMSD quantifies shape similarity of buildings based on their true 3D boundaries without simplification or projection. SPHARM RMSD is basically consistent with *volume* accuracy metrics, albeit with a larger value range and standard deviation. The normalized RMSD provides a metric comparable to the other two components.

The examples given in Figure 5-7 is visually consistent with the shape similarity rates reported in Table 5-2. In the first two example buildings in (a) - (d), the shape similarity is now at 0.55. For the Alumni Hall in (a) and (b), the extracted building fails to extract the sector-shape in the front of the building. Furthermore, the rain canopy in front of the building is not counted in the reference building, whereas it is detected as a part of the building in the reference, which leads to even lower shape similarity. For the Middlesex College, the major problem leads to the dissimilarity is the steeple of the building, which is shorter and wider in the extracted building. As SPHARM has to expand buildings with low compactness into a unit sphere, worse shape similarity tends to result for such buildings. The latter two buildings, Ivey School and Law School in (e) – (h) have high shape similarity near 0.85. They both have corresponding parts between reference and extracted buildings. Moreover, their shape has higher compactness, and is easier to transform to a unit sphere. In summary, the visual similarity for buildings is consistent with the rate given by SPHARM. For complicated buildings such as Middlesex College, it is better to separate the buildings into parts and compare each highly compacted part correspondingly to avoid sensitivity during the SPHARM deformation.

Corner point distance RMSD is around 1m in each direction between reference buildings and reconstructed buildings. The correlation analysis of the metrics in Table 5-2 demonstrates that the only significant correlation (coefficient 0.77, p-value < 0.05) is between DX and DY. That is because when a corner point is shifted from reference building to sample building, the shift usually happens on X and Y direction simultaneously. As shown in Figure 5-6, the positional differential vectors visually explain the residual distribution. The distribution map can provide valuable information
for reconstruction method improvement, as well as an accuracy baseline for future application comparisons.

5.5 Conclusion
In this Chapter, the evaluation method is decomposed as three stages: reference data, evaluation features, and comparison function. A detailed and accurate reference dataset is critical to provide a benchmark for the evaluation process. The reference building usually has a higher level of detail than the reconstructed building. Based on the reference data, evaluation features are designed to deliver building characteristics that can distinguish reference building from sample building. A good evaluation feature is usually robust, straightforward, and self-explained. Finally, comparison functions quantify the difference in evaluation features.

To highlight the difference between 3D to 2D building evaluation, this Chapter proposes a multi-criteria 3D building evaluation system. The proposed evaluation system is based on three components: volume, surface, and point. The volume accuracy reports the percent of correctly reconstructed building volume, which expresses the chance of occurrence of either missing or excess parts in the final reconstruction. The surface accuracy matches 3D surface directly between reference and reconstructed buildings. Surface comparison including rooftops and walls is implemented by the sphere parameterization, followed by the SPHARM expansion. This surface comparison method is essentially different from existing 3D evaluation methods that only take rooftops or matching building surfaces facet-by-facet into account during evaluation. The point accuracy provides positional accuracy for reconstructed buildings at feature points, such as corners and centroids.

The proposed multi-criteria evaluation system is tested by a simple LiDAR-based 3D building reconstruction. The resulting metrics from example buildings show that volume accuracy has less room for improvement with smaller standard deviation, in comparison with the other two components. These issues restrict volume accuracy to evaluate further improved reconstruction independently. The surface and point accuracy metrics provide
supplementary assessment from different perspectives, with most metrics reporting low and insignificant correlations. The surface comparison conducted in a true 3D environment makes this evaluation system stand out from current evaluation methods that simplified or projected the 3D building for evaluation. Together, these three components provide an improved system for evaluation of an entire building. This improved evaluation system is expected to contribute positively to the assessment of future 3D buildings using advanced reconstruction methods and high-resolution terrestrial data.

References


Chapter 6: General discussion and conclusions

6.1 Summary
This thesis deals with automatic extraction of building information from remotely sensed data. Automatic information extraction has always been a popular topic. The experience and algorithms developed for buildings from this thesis can potentially be applied to the extraction of other man-made and natural ground features. Effective and stable building information extraction methods save time and labour costs to provide a valuable database for other applications. As additional new methods are developed, their capacity to provide accurate building information is critical for method comparison and new method development.

Methodologies for extracting a building’s information can range from simple to sophisticated, from 2D to 3D. Using characteristics to separate buildings from its surrounding environment (i.e., colour or height) help identify a building in an image. The extracted building boundary is helpful in extracting building height. With the inclusion of additional advanced and high-resolution images, rooftop details can be reconstructed. Accuracy is an important factor for each information extraction step; thus, accuracy at each stage is discussed.

Chapter 2 presents a multi-criteria and hierarchical evaluation system for building extraction from remotely sensed data. Most of current evaluation methods are focused on classification accuracy, while the other dimensions of extraction accuracy are usually ignored. The proposed evaluation system consists of three components: (1) the matched rate, including evaluation metrics for the traditional classification accuracy (e.g., Completeness, Correctness and Quality); (2) the shape similarity that describes the resemblance between reference and extracted buildings, including image-based and polygon-based metrics; and (3), the positional accuracy which is measured by distances between reference and extracted buildings at feature points such as a building’s centroids. The system also hierarchically evaluates extracted buildings at per-building, per-scene, and overall levels. To reduce redundancy among different metrics, principal component analysis and correlation analysis are employed for metrics selection and aggregation. Four different building extraction
methods, using high-resolution optical imagery and/or LiDAR data, are implemented to test the proposed system. This system can highlight perceptible differences between the extracted building footprints and the reference data, even if this difference is insignificant as measured by traditional metrics.

Furthermore, in a digital surface model (DSM) derived from high-resolution stereo images, the accuracy of building rooftop elevations is usually much lower than that of open areas without tall objects. This problem makes building height estimation difficult from the stereo images. Inaccuracy in building rooftop elevation is caused by many factors including shadows, occlusions, smoothing constraints in the matching algorithms, and the mismatch on extremely tall buildings. In order to improve the accuracy of building rooftop elevations, existing image matching methods are effectively improved by adding building footprint maps as a constraint, as demonstrated in Chapter 3. The proposed image matching method consists of three steps. Initially, a building footprint is used to identify corresponding building rooftop locations in the stereo images. Secondly, the left and right stereo images are matched according to colour and shape. Thirdly, the left and right sub-images of the building are further matched at pixel level to generate detailed roof elevations. Validation using surveyed rooftop elevations demonstrates that the proposed method can estimate building rooftop elevation with one third of the error typically generated using the current commercial software. In addition, unlike the results from current software, the errors for low rise and high rise buildings are consistent using the proposed method.

Chapter 4 presents a building height estimation method from digital surface models (DSM). The DSM derived from a single pair of optical stereo images is affected by occlusions and shadows, which leads to indistinct building borders in the DSM. To extract building height from such DSM with the assistance of building footprints, a “building-ground elevation difference model” (EDM) has been designed in this study. This model describes the trend of elevation difference between a building and its neighbours, in order to find a stable elevation difference. This stable difference is used to compute building’s height. The EDM is discussed under both flat and sloped ground situations. Experiments on two study sites using the proposed model demonstrate that estimated height at rooftop points obtained accuracies within 1.5m, which out-perform the conventional filtering method. Exploring the capacities
of the proposed model, three types of buildings are reconstructed by clustering their height histograms.

Finally, Chapter 5 presents a multi-criteria system to evaluate the accuracy of reconstructed 3D buildings. Current 3D evaluation methods are derived from 2D pixel-based evaluation; however, the difference between 2D and 3D evaluation methods has not been well presented in previous literature. Most 3D building evaluation methods concentrate solely on rooftop accuracy, while ignoring the degree of accuracy found with regards to walls. To address these problems, in this chapter a multi-criteria evaluation system is designed based on three components: volume, surface, and point. The volume accuracy component represents the traditional classification accuracy based on random samples. The surface accuracy component evaluates shape similarity, which compares the sample and reference buildings, including rooftops and walls, in a true 3D environment. The point accuracy component measures the distance at feature points between a sample building and a reference building. This multi-criteria system aims to provide an improved evaluation method for building reconstruction using advanced algorithms and multi-platform data. The system also expects to provide valuable information to guide applications with different accuracy requirements.

Automated information extraction from images, and more broadly, image understanding, has been a popular research topic in the last few decades. It is still a difficult task for computers to understand images with one hundred percent accuracy. This difficulty is due to the image complexity and the target variety. Buildings in urban areas vary greatly in their colour, shape, size, height, and structure. This thesis is one of the many efforts to enhance the capability of computer algorithms to understand the Earth surface features closer to human vision.

6.2 Conclusions and Contributions

This thesis has accomplished four research objectives presented at the end of Introduction Section 1.5. The achievements of this thesis for extracting and evaluating building information are described as follows.

- Two-dimensional building extraction and evaluation (2D)
Four building extraction methods are described and implemented. The four methods use different data sources, including the original colour and infrared-red (CIR) images, LiDAR data and their combination. The last method combines the LiDAR data with a CIR image after relief displacement correction. Building boundaries are detected by identifying discontinuity at building edges; this discontinuity can be colour, height, texture, or other variants that can distinguish a building from its surrounding environment.

A comprehensive evaluation method is innovatively designed to assess 2D building extraction accuracy. The designed system demonstrates its superiority to traditional evaluation methods, especially when traditional classification accuracy cannot distinguish between the performances of two extraction methods, while the proposed system can identify the best performer very well. Finally, the impact of building types on evaluation methods is also analyzed by taking advantage of this multi-level system.

- **Building height estimation (2.5D)**

Building height has been estimated from two different perspectives: directly from stereo image matching or from DSM. Regarding direct image matching and height estimation, in this thesis an advanced building height estimation method has been developed from stereo imagery constrained by building footprints. Focusing on the bottleneck of current matching methods which perform poorly in both dense urban areas, and on tall buildings, this method is novel in the sense that it uses building footprints as a constraint to enhance image matching.

Compared with DSMs generated from popular commercial software, the DSM created in this study is demonstrably superior for extracting building rooftop elevation with high accuracy. Experiments have been carried out with GeoEye-1 stereo images in Beijing using the proposed method and two commercial stereo image matching modules (ENVI and PCI). It has improved the DSM accuracy over tall buildings’ roofs, with the elevation errors for tall buildings reduced by 90% in comparison with current commercial software. Whereas the DSMs derived using the commercial software show difference in rooftop
elevation errors between low and tall buildings, the proposed matching method keeps elevation accuracy for low and tall buildings consistent. The comparison between a direct SGM method and this study also indicates that the proposed method can generate more accurate measures of building elevation than the SGM algorithm.

Considering building height estimation from inaccurate DSMs, this thesis designed an estimation method for building height from previously derived DSMs. This height estimation method, referred to as an “elevation difference model (EDM),” works in dense urban areas without any ground assumptions (such as widely used “flat ground”). It is a reliable method for processing inaccurate DSMs over flat and sloped terrain and in the complicated urban scenes.

Comparing the results of the original EDM with the slope-adjusted EDM, slope adjustment can remove the underestimation of bare-ground elevation; thus reducing the overestimation of the building height. The EDM reaches similar levels of accuracy in both open, spacious areas and high-density urban areas. This is due to the self-adaptive nature of the EDM that allows it to locate the distance where the buffering effect of building elevations on neighbours can be removed and the ground elevation can be found.

From the comparison of the proposed EDM and other height extraction methods, the results show that the slope-adjusted EDM is more accurate than the DSM filter approach. Even without an extra DTM, the slope-adjusted EDM method produced comparable or better accuracy to the independent DTM method. The height error of the testing points is less than 1.5m.

- Three-dimensional building reconstruction and evaluation (3D)
  Three-dimensional buildings are reconstructed from LiDAR data and a multi-criteria system is developed to evaluate the 3D reconstruction accuracy from different perspectives. This novel 3D building evaluation method assesses 3D building surfaces, including roofs and walls, in a simple and consistent way. It stands out from all other similar methods in previous literature because it is implemented in a true 3D environment
and compares all faces of a building as a whole, rather than face-to-face in a 2D environment.

The designed multi-criteria evaluation system is tested by a simple LiDAR-based 3D building reconstruction. The resulting metrics, including the volume, surface and point components, from example buildings show that volume accuracy has less room for improvement with smaller standard deviation, in comparison with the other two components. These issues restrict the use of volume accuracy to evaluate further improved reconstruction. The surface and point accuracy metrics provide supplementary assessment from different perspectives, with most metrics reporting low correlations. The surface comparison conducted in a true 3D environment makes this evaluation system stand out from current evaluation methods that simplified or projected the 3D building for evaluation.

6.3 Future research
With the demonstrated contributions, this work reveals numerous areas for further investigation. Future works based on this thesis are expected to be built on the availability of ever higher image (spatial, temporal, angel) resolutions, to develop smarter algorithms and retrieve more accurate building information.

6.2.1 Very high-resolution images for 2D building footprint extraction
In the coming decade, the advent of new, more advanced sensors will enhance the current remote sensing earth observation capacities. With the development of unmanned aerial vehicle (UAV) techniques (Herwitz et al., 2004) and the future launch of very high-resolution satellite sensors such as the WorldView-3 satellite (DigitalGlobe, 2013), the primary data for building information extraction will be optical imagery with high image overlap rate and spatial resolution at the centimeter level. Optical images include colour and edge information which is useful to extract building boundaries directly and they carry height information which can be used to extract buildings indirectly (via stereo
image matching). Both edge and colour discontinuities can be employed to search for 2D building outlines.

Very high spatial resolution images bring opportunity and challenge to building information extraction at the same time. On one hand, the centimeter level spatial resolution images improve their capability to detect ground objects. Details of ground objects can be perceived easily in very high spatial resolution imagery. Textures of ground objects can also be employed to distinguish objects. Moreover, the higher spectral resolution and wider spectral response range provide further information to separate ground objects based on their spectral signatures. On the other hand, an increase of spatial resolution in imagery is accompanied by disadvantages. Data volume for the same area will increase dramatically due to improvements in spatial resolution. Very high resolution images usually taken by smaller and more sensitive sensor units more readily generate electronic noise (Clark, 2013) and affect the image quality.

In addition, the spatial resolution improvement will be accompanied by a scale issue. For instance, in the 1980s, Landsat images with coarse resolutions required pixel unmixing techniques in order to identify ground features (Petrou & Foschi, 1999) because the sizes of the ground objects are usually smaller than an image pixel area. In the 1990s, improvements in image spatial resolution made pixel size comparable to object size, thus pixel-based classification has been widely used for land use and land cover classifications (Yang et al., 2010). In the twenty-first century, the advent of sub-meter satellite sensors propels the object-based segmentation and classifications, since the pixel size is smaller than the ground object size in this stage. In 2010, continuing improvements on sensor ability produces images at the centimeter resolution level. At the same time, high-performance computers make sophisticated and computationally intensive algorithms feasible. A new trend is to process images and extract information hierarchically. Specifically in building information extraction, an image pyramid is built based on very high-resolution images. Building location and height are estimated on downscaled images first, and further details are enriched with the increase of image resolution based on the image pyramid.
6.2.2 Multi-view stereo image matching and multi-platform image integration

Currently, buildings and other urban features are observed using images from different platforms. The space-borne platform provides large-scale imagery at sub-meter spatial resolution. The airborne platform is the major data source used to detect buildings from multi-views with redundant information. The terrestrial platform provides a unique capacity to obtain building wall details. The integration of images from these various platforms is expected to provide more information about buildings and to better detect building information.

An accurate building footprint extraction method requires separating buildings from other impervious surfaces using DSM. Multi-view optical imagery matching using semi-global matching (Hirschmuller, 2008) is claimed to be able to generate DSM at centimeter levels, which is comparable to LiDAR derived DSM. If a series of very high-resolution aerial photos can generate accurate building height information, optical imagery itself can be used to extract building footprint with direct spectral (colour) and indirect height information.

To further improve the building footprint extraction, the advantage of multi-platform data needs to be maximized. Image registration between different platforms (Zouqi, 2013) can construct a seamless colour information network. Feature detection operators, such as the scale-invariant feature transform (SIFT) operator (Lowe, 2004), are useful tools for detecting feature points and connecting different images from different platforms. For most buildings, edge discontinuity is obvious with regard to both colour and height properties. Thus, methods using a combination of edge and height to maximize the discontinuity detection is expected to find building edges at sub-pixel accuracy.
Figure 6-1. The illustration of multi-platform image integration
6.2.3 Combining terrestrial sensors with air-borne sensors for 3D building reconstruction

Aerial imagery and LiDAR data have been used to reconstruct building roofs; however, such data lacks the ability to reconstruct wall facades, even when oblique view images are used (Petrie, 2009). Recent terrestrial data, such as terrestrial LiDAR data (Pu & Vosselman, 2009), handheld digital cameras (Bhatla et al., 2012), or video streams (Brilakis et al., 2011; Pollefeys et al., 2008), can act as complementary data. Their use indicates a new trend for 3D building reconstruction.

In the future, 3D buildings may be reconstructed based on both aerial photos and terrestrial hemispherical (fisheye) images. Roof details may be extracted using aerial photos through stereo image matching and photogrammetry techniques, while the wall details may be generated by image processing over terrestrial imagery. A reconstructed 3D building model with both rooftop parts and wall details is complete and comparable to the real world case. In the process of automatic 3D building reconstruction, some technical issues need to be solved. For example, the edges that separate a roof and the walls, referred to as “eaves,” are required to be identified in both aerial photos and terrestrial images. For complicated wall structures (i.e., with columns), transforming the description of a wall feature into a data structure will be difficult. Additionally, distortion within images may occur due to the use of the hemispherical camera lens and the perspective view effect.

6.2.4 Building information application in natural disaster management

Building information provides essential information to many applications such as natural disaster management. With accurately extracted 2D and 3D building information, building vulnerability during a disaster has been modeled with building information as input (Geiß & Taubenböck, 2012). Earthquake hazard analysis can be divided into pre-event and post-event analysis. Pre-event studies focus on reduction (mitigation) and readiness (preparedness), while post-event research concentrates on emergency response, environmental impact assessment and post-event recovery.
The post-event applications include such efforts as building damage maps for rapid response and rescue works. With images before and after natural disasters, collapsed and partially damaged buildings can be identified by their colour and height change (Corbane et al., 2011; Hussain et al., 2011). With post-event images, this type of study is scientifically reasonable and can be validated (Dong & Shan, 2013). In the pre-event research, however, it is difficult to report a building vulnerability map in a reasonable manner because there is no actual earthquake occurring to test the map accuracy (Mück et al., 2013; Wu et al., 2013).

With the 2D and 3D building information, building vulnerability to natural disasters can be modelled. For example, a building’s wall material and structure are critical foundations of a building’s ability to resist earthquakes (FEMA, 2010). According to the European macroseismic scale (Grunthal, 1998), buildings made of masonry, wood, or concrete have different empirical collapse fragility functions (Jaiswal et al., 2011). Equally, reinforced structures and sheared-walls can improve a building’s resistance to earthquakes (FEMA, 2010). A study of Vancouver and southern British Columbia also reported building types and their respective empirical fragile curves according to an advisory panel of 71 specialists (Ventura et al., 2005). Although these variables cannot be extracted directly from remote sensing techniques, building structure and material can be inferred by its location, age, height, and exterior style. Furthermore, other building variables can be extracted from remote sensing imagery, such as building size (Mück et al., 2013), height (Meslem et al., 2012), shape (Pimanmas et al., 2010), soil condition (Harp & Jibson, 2002), and regolith depth (Shafique et al., 2011).

Efforts have been made to integrate different variables. Mueller et al. (2006) used building information, geology and soil condition, and building context information in a vulnerability analysis. This preliminary study with simple variables demonstrates that building-related variables can be derived from remotely sensed data either directly or indirectly. This work is based on the recognition of the urban building structure types. Studying a coastal city, Mück et al. (2013) investigated building vulnerability in an earthquake together with tsunami. Many physical and social variables, such as building height, material, structure, and hammer test value, are integrated and ranked to classify buildings into different levels. In another study, a
reference Social Vulnerability Index (SVI) was created from census data (Ebert et al., 2009). By evaluating proxy-variables, a stepwise regression model was applied to select the best explanatory variables for changes in the SVI. Finally, natural disaster analysis models, such as HAZUS (FEMA, 2010), GEM Openquake, and Riskscape, provide frameworks for variable integration and geo-visualization.

In summary, the application of building information for natural disaster management is still a challenge. Much work can still be done to improve the building vulnerability prediction model accuracy by cooperating with earthquake experts.

References


Appendix A: Traditional accuracy assessment for image classification

Image classification is usually the prerequisite for thematic mapping. It is also an important transition from image DN values to useful information. As illustrated in Figure A-1, a given image has wide range of DN value, is classified into four classes.

It is important to report the accuracy of the image classification. Ground surveyed data or other reference data are usually collected to evaluate the thematic map accuracy. The confusion matrix is the core part of the accuracy assessment (Foody, 2002) in traditional image classification. An example of confusion matrix is provided in Figure A-2, where a matrix reports a simple cross-tabulation of the mapped class label against that which is observed in the ground or reference data for a sample of cases at specified locations. Confusion matrices provide the basis on which to both describe classification accuracy and characterize errors, which may help refine the classification.

Based on the confusion matrix, common measures of classification accuracy include percentage correctness, the user’s accuracy, the producer’s accuracy, and the Kappa coefficient. The highlighted elements (grey cells in left diagram of Figure A-2) represent the main diagonal of the matrix that contains the cases, where the class labels depicted in the image classification and ground data set agree, whereas the off-diagonal elements contain those cases where there is a disagreement in the labels.
Figure A-1. The process of thematic mapping. From image and its DN values to classes, data volume is compressed.
Figure A-2. Confusion matrix and traditional accuracy measures. Adapted from Foody (2002)

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>n_{AA}</td>
<td>n_{AB}</td>
<td>n_{AC}</td>
<td>n_{AD}</td>
<td>n_{A+}</td>
</tr>
<tr>
<td>B</td>
<td>n_{BA}</td>
<td>n_{BB}</td>
<td>n_{BC}</td>
<td>n_{BD}</td>
<td>n_{B+}</td>
</tr>
<tr>
<td>C</td>
<td>n_{CA}</td>
<td>n_{CB}</td>
<td>n_{CC}</td>
<td>n_{CD}</td>
<td>n_{C+}</td>
</tr>
<tr>
<td>D</td>
<td>n_{DA}</td>
<td>n_{DB}</td>
<td>n_{DC}</td>
<td>n_{DD}</td>
<td>n_{D+}</td>
</tr>
<tr>
<td>Σ</td>
<td>n_{+A}</td>
<td>n_{+B}</td>
<td>n_{+C}</td>
<td>n_{+D}</td>
<td>n</td>
</tr>
</tbody>
</table>

Percentage correct = \(\frac{\sum_{k=1}^{q} n_{kk}}{n} \times 100\)

User’s accuracy = \(\frac{n_{ii}}{n_{i+}}\)

Producer’s accuracy = \(\frac{n_{ii}}{n_{+i}}\)

Kappa coefficient =
\[
\frac{n \sum_{k=1}^{q} n_{kk} - \sum_{k=1}^{q} n_{k+} \cdot n_{+k}}{n^2 - \sum_{k=1}^{q} n_{k+} \cdot n_{+k}}
\]
Appendix B: Object-based classification and nearest neighbour classifier

As data spatial resolution increases, the relationship between image cell size and ground feature dimensions is largely changed. This change is illustrated in Figure B-1. In order to remove image noise and conduct image processing effectively, a technique has been developed to partition remote sensing imagery into meaningful image-objects, thereby allowing assessment of their characteristics through spatial, spectral and temporal scales. This technique is called object-based image analysis (OBIA), or geospatial object based image analysis (GEOBIA) in cases that involve geospatial information (Blaschke, 2010).

The most common approach used to convert pixels into objects is image segmentation, which is generally divided into four categories: (a) point-based, (b) edge-based, (c) region-based and (d) combined, from an algorithmic perspective (Blaschke, 2010). Segments are regions generated by one or more criteria of homogeneity in one or more dimensions (of a feature space) respectively. Thus segments have additional spectral information compared to single pixels, but of even greater advantage than the diversification of spectral value descriptions of objects is the additional spatial information for objects. It has been frequently claimed that this spatial dimension (distances, neighbourhood, topologies, etc.) is crucial to OBIA methods.

After image segmentation and readied image objects, object-based classification is the next step in converting images into useful maps with classes. The Nearest Neighbour classifier is a widely used classification method, by selecting a set of samples of different classes to assign membership values. The procedure consists of two major steps (Trimble, 2011): (a) Teaching the system by giving it certain image objects as samples. (b) Classifying image objects in the image object domain based on their nearest sample neighbours. The Nearest Neighbour classifier returns a membership value of between zero and one, based on the image object’s feature space distance to its nearest neighbour. The membership value has a value of one if the image object is identical to a sample. If the image object differs from the sample, the feature space distance has a fuzzy dependency on the feature space distance to the nearest sample of a class, as illustrated in Figure B-2. The user can select the features to be considered for the feature space.
Figure B-1. Relationship between objects under consideration and spatial resolution. This figure is adapted from Blaschke (2010), where (a) low resolution: pixels significantly larger than objects, sub-pixel techniques needed. (b) medium resolution: pixel and objects sizes are of the same order, pixel-by-pixel techniques are appropriate. (c) high resolution: pixels are significantly smaller than object, regionalization of pixels into groups of pixels and finally objects is needed.
For an image object to be classified, only the nearest sample is used to evaluate its membership value. The effective membership function at each point in the feature space is a combination of fuzzy function over all the samples of that class. When the membership function is described as one-dimensional, this means it is related to one feature, as illustrated in Figure B-3.

In higher dimensions, it is harder to depict the membership functions depending on the number of features considered. However, if you consider two features and two classes only, it might look like the graph on Figure B-4.

The following images are an example to demonstrate about how images are segmented, samples are selected, and objects are classified in software Ecognition (Trimble, 2011).
Figure B-2. Membership function created by Nearest Neighbour classifier. Adapted from Trimble (2011).
Figure B-3. Membership function showing Class Assignment in one dimension. Adapted from Trimble (2011)
Figure B-4. Membership function showing Class Assignment in two dimensions. Samples are represented by small circles. Membership values to red and blue classes correspond to shading in the respective colour, whereby in areas in which object will be classified red, the blue membership value is ignored, and vice-versa. Note that in areas where all membership values are below a defined threshold (0.1 by default), image objects are not classified; those areas are coloured white in the graph. Adapted from Trimble (2011)
Figure B-5. Multi-resolution segmentation on aerial imagery with 3 bands. It is a pseudo-colour image with green, red, and near infrared bands of 0.3m spatial resolution on the University of Western Ontario campus, London, Canada. Two criteria, shape and colour, are used for image segmentation.
Figure B-6. Class definition and sample selection. There are five classes defined: bare soil, impervious surface, shadow, vegetation, and water. Samples for each class are selected based on the three spectral bands and the NDVI layer of an aerial image.
Figure B-7. The result of Nearest Neighbor classification based on the samples. Blue represents water; green represents vegetation; yellow represents soil; red represents buildings; and cyan represents other impervious surfaces. Buildings and other impervious surfaces are distinguished by their shape parameters.
Appendix C: LiDAR data collection, processing, and related software packages

Light Detection and Ranging (LiDAR) is an active remote sensing technology that measures distance by illuminating a target with a laser and analyzing the reflected light (Lillesand et al., 2008). It is a method that uses light in the form of a pulsed laser to measure ranges (variable distances) to the Earth. These light pulses — combined with other data recorded by the airborne system — generate precise, three-dimensional information about the shape of the Earth and its surface characteristics.

A LiDAR instrument principally consists of a laser, a scanner, and a specialized GPS receiver. Airplanes and helicopters are the most commonly used platforms for acquiring LiDAR data over broad areas (NOAA, 2013). Topographic LiDAR typically uses a near-infrared (about 1040nm) laser to map the land. LiDAR systems are applied to examine both natural and manmade environments with accuracy, precision, and flexibility. Such applications include accurate shoreline measurement, digital elevation models generation, emergency response, etc.

LiDAR point clouds carry height and intensity information. Taking the advantage of an active sensing system, LiDAR can measure the distance from a sensor to the ground precisely based on the signal travel time. With a highly accurate clock, a GPS, and an Inertial Measuring Unit (IMU), LiDAR can map the ground terrain even in steep slopes and shadowed areas. In addition, LiDAR also measures the backscattered energy from the target (Yan et al., 2012). The strength of returned energy from each pixel is based on the Earth surface characteristics, and the corresponding measurements form the intensity map. For example, water absorbs LiDAR waves and returns little or no energy, while certain roof types have strong reflection. Intensity is used as an aid in feature detection and extraction, in LiDAR point classification, and as a substitute for aerial imagery when none is available.

Furthermore, modern LiDAR systems are able to record up to five returns per pulse, which demonstrates the ability of LiDAR to discriminate between not only such features as a forest canopy and bare ground but also surfaces with a range of covers (Lillesand et al., 2008). In urban areas, the first return of LiDAR data typically measures the elevations of tree canopies, building roofs, and other unobstructed surfaces. The second and later returns, usually applied for vegetation, record the return signal after penetrating the canopies. The last return reflects the bare ground surface, as LiDAR near-infrared waves cannot penetrate the ground.
Figure C-1. The illustration of LiDAR imaging system. Laser beam may generate several reflections, typically multi-pulses on a tree. Picture courtesy of (http://www.imagingnotes.comgoarticle_freeJ.phpmp_id=264)
Software and expertise for defining the first surface and the ground are delivering proven results. This LiDAR application is particularly well suited for the generation of digital DEMs, topographic contouring, and automatic feature extraction. For instance, LiDAR data can be utilized for ground information extraction, especially detect the shape and height of above-ground objects (Chen et al., 2007; Lee & Younan, 2003). A generally used method is filtering based on morphological operations with various algorithms designed to keep the terrain features unchanged while using large window sizes for the morphological opening, as illustrated in Figure C-3. There are many commercial and open-source software packages for LiDAR point cloud processing, as described in Table C-1.

As an example of LiDAR processing software, the “LiDAR Analyst for ArcGIS” software is used to process LiDAR point cloud and detect building footprints. These airborne LiDAR point clouds were acquired with Optech ALTM 3100 sensor. The data were collected on May 20, 2006 during leaf-off conditions in London Ontario, allowing for maximum penetration of LiDAR pulse to the ground. This sensor is capable of recording the first, second, third and last pulse returns, as well as intensity data.
Figure C-2. An example of LiDAR image. (a) A LiDAR range (height) image. (b) The corresponding LiDAR intensity image.
Figure C-3. Example of one-dimensional laser points. Adapted from Chen et al. (2007): (a) measured points, (b) points whose elevations are updated by opening with a neighbourhood of 3, (c) points whose elevations are updated by opening with a neighbourhood of 7, (d) the point-wise difference between neighbourhood 3 and 7, and (e) the updated points after one iteration.
Table C-1. A summary of some popular LiDAR processing software packages

<table>
<thead>
<tr>
<th>Software Packages</th>
<th>Description</th>
<th>Open source</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP360</td>
<td>LP360 for ArcGIS is an extension to ArcMap as well as a standalone version for Windows that allows visualizing and processing of very large point clouds. Available in three levels of capability, LP360 provides tools from rapid visualization and derived product generation through advanced features such as automatic ground classification and building footprint extraction.</td>
<td>No</td>
</tr>
<tr>
<td>LiDAR Analyst</td>
<td>LIDAR Analyst was initially developed and released by Visual Learning Systems in 2005 and now owned by Overwatch. LIDAR Analyst is an automatic feature extraction application that uses airborne LIDAR (Light Detection and Ranging) data to create three-dimensional vector objects</td>
<td>No</td>
</tr>
<tr>
<td>SCOP++</td>
<td>SCOP++ software provides fast and fully automatic classification of LiDAR point clouds into terrain and off-terrain points. The new enhanced filtering process now easily detects large man-made structures as well as low vegetation objects.</td>
<td>No</td>
</tr>
<tr>
<td>VG4D</td>
<td>VG4D SmartLiDAR solution is a complete end-to-end, standalone software solution for all different types of LiDAR/Point Cloud datasets. It then uses that accurate dataset to extract vital information through an optimized streamlined workflow.</td>
<td>No</td>
</tr>
<tr>
<td>LASTools</td>
<td>LASTools is a collection of command line tools to classify, tile, convert, filter, raster, triangulate, contour, clip, and polygonize LiDAR data.</td>
<td>Partial</td>
</tr>
<tr>
<td>BCAL LiDAR Tools</td>
<td>BCAL LiDAR Tools are open-source tools developed by Idaho State University, Boise Centre Aerospace Laboratory (BCAL). These tools can be used for processing, analyzing, and visualizing LiDAR data.</td>
<td>Yes</td>
</tr>
<tr>
<td>FUSION /LDV</td>
<td>FUSION is a LiDAR data conversion, analysis, and display software suite. FUSION allows 3-dimensional terrain and canopy surface models and LiDAR data to be fused with more traditional 2-dimensional imagery. FUSION processes raw LiDAR data into a number of vegetation metrics. Canopy-and ground-level surface models can be produced.</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Figure C-4. The study area is cropped from a LiDAR point cloud, consisting of many LiDAR stripes
Figure C-5. Resulting rasterized image after interpolation from a LiDAR point cloud
Figure C-6. The extracted bare ground layer by removing above ground objects.
Figure C-7. The normalized digital surface showing only above ground objects.
Figure C-8. Extracted building boundaries based on height, shape and LiDAR return signals.
Appendix D: Principal component analysis (PCA) and factor selection

Principal components analysis (PCA) is a technique that transforms the original remotely sensed dataset into a substantially smaller and more readily-interpreted set of uncorrelated variables that represent most of the information present in the original dataset (Jensen, 2005). Originally proposed in (Wold et al., 1987), PCA has been used for data compression and analysis in multispectral and hyper-spectral remote sensing. Its aim is to reduce a larger set of variables (or data dimension) into a smaller set of 'representative' variables, called 'principal components', which account for most of the variance in the original variables. As in the example provided in Figure D-1, two dimensional temperature and pressure data can be represented by a coordinate system \( v_1 \) and \( v_2 \), where the majority of information can be described merely by the first component \( v_1 \).

Mathematically, PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance is attributed to the second coordinate, and so on.
Figure D-1. An example of principal components analysis. Image courtesy of (http://www.mech.uq.edu.au/courses/mech4710/pca/s1.htm)
To compute each component specifically for remote sensing image bands, the covariance matrix (Cov) for all involved bands is computed. The eigenvalues, $E=\{\lambda_{1,1}, \lambda_{2,2}, \lambda_{3,3}, \ldots, \lambda_{n,n}\}$, and eigenvectors $EV=[a_{kp}\ldots$ for $k=1$ to $n$ bands, and $p=1$ to $n$ components], of the covariance matrix are computed such that:

$$EV \cdot Cov \cdot EV^T = \begin{bmatrix}
\lambda_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \lambda_{2,2} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \lambda_{3,3} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \lambda_{4,4} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \lambda_{5,5} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \lambda_{6,6} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \lambda_{n,n}
\end{bmatrix}$$

where $\lambda_{ii}$ in the diagonal covariance matrix $E$ is the variance for $1$st to $n$th principle components. With the eigenvector, each principle component can be calculated by summarized each old band with given weight:

$$DN'_i = \sum^n_k (a_{ki} \times DN_k), \quad k = 1, \ldots, n$$  \hspace{1cm} (A-1)

Where $DN$ is the old bands, $DN'$ is the new principle components. In addition, with $\lambda_{ii}$ available, the weight for each band (“loads”) associated with $i$th principle component can be calculated as follows:

$$w_{ki} = \frac{a_{ki} \times \sqrt{\lambda_i}}{\sqrt{Var_k}}$$  \hspace{1cm} (A-2)

where $a_{ki}$ is eigenvector for band $k$ and component $i$, $Var_k$ is the variance of band $k$ in the covariance matrix. $w_{ki}$ describes the correlation for components and bands (Jensen, 2005), thus can contribute to the factor selection.
Appendix E: Ground GPS survey forms and pictures

The ground survey is performed to measure building footprint, height, and elevation using GPS and laser range finder. As reference data, this information is used to evaluate the extracted building information.

Figure E -1. Laser rangefinder to measure building height. The surveyor on the right was measuring a building’s height based on its base and top distance and angle. The surveyor on the left was recording the height and take regular photos for this building.
Figure E-2. Differential GPS station installed on the roof top of a tall building
Figure E-3. Differential GPS receiver to measure building elevations (a) on roofs and (b) at corners
Figure E-4. An example of a record sheet with each building’s name, relative heights, base GPS points and photo number and direction
Appendix F: Semi-variogram model

The Earth's surface and remotely sensed imagery contain spatial information that, if quantified, could be used to optimize many sampling procedures in remote sensing (Curran, 1988). The semi-variogram is a function that relates semi-variance to sampling lag. This function can be estimated using remotely sensed data or ground data and represented as a plot that gives a picture of the spatial dependence of each point on its neighbour. In the early age of remote sensing, semi-variogram has been widely used for the selection of the most appropriate spatial resolution (Webster & Oliver, 2001; Woodcock & Strahler, 1987). An example of a typical semi-variogram is given in Figure F-1, while the explanation of corresponding terms is given in Table F-1.

Mathematically, the theoretical variogram $2\gamma(x,y)$ is a function describing the degree of spatial dependence of a spatial random field or stochastic process $Z(x)$. It is defined as the variance of the difference between field values at two locations (x and y) across realizations of the field (Curran, 1988):

$$2\gamma(x, y) = \text{var}(Z(x) - Z(y)) = E\left(\left|\left(Z(x) - \mu(x)\right) - \left(Z(y) - \mu(y)\right)\right|^2\right). \quad (A-3)$$

If the spatial random field has a constant mean $\mu$, this is equivalent to the expectation for the squared increment of values between locations x and y (where x and y are not coordinates, but points in space):

$$2\gamma(x, y) = \text{var}(Z(x) - Z(y)) = E\left(\left|\left(Z(x) - \mu(x)\right) - \left(Z(y) - \mu(y)\right)\right|^2\right); \quad (A-4)$$

where $\gamma(x,y)$ itself is called the semi-variogram.

The empirical semi-variogram and covariance provide information on the spatial autocorrelation of datasets. In order to organize the data samples and search the underneath autocorrelation rules, parametrically pre-defined empirical models are generally used to fit the sample data distribution.
Figure F-1. A typical example of semi-variogram
Table F-1. The terms and symbols used in the description of the semi-variogram. Adapted from Curran (1988)

<table>
<thead>
<tr>
<th>TERM</th>
<th>SYMBOL</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support</td>
<td>-</td>
<td>Area and shape of surface represented by each sample point.</td>
</tr>
<tr>
<td>Lag</td>
<td>h</td>
<td>Distance (and direction in two or more directions) between sampling pairs</td>
</tr>
<tr>
<td>Sill</td>
<td>s</td>
<td>Maximum level of $\gamma$ (h).</td>
</tr>
<tr>
<td>Range</td>
<td>a</td>
<td>Point on h axis where $\gamma$ (h) reaches maximum. In sample data where $\gamma$ (h) reaches approximately 95% of the sill. Places closer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>than the range axe related, places further apart are not.</td>
</tr>
<tr>
<td>Nugget variance</td>
<td>C0</td>
<td>Point where extrapolated relationship $\gamma( h)/h$ intercepts the $\gamma$ (h) axis. Represents spatially independent variance.</td>
</tr>
<tr>
<td>Spatially dependent</td>
<td>C</td>
<td>Sill minus nugget variance.</td>
</tr>
<tr>
<td>structural variance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table F-2. Some popular empirical semi-variogram models

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Semi-variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\gamma_e(h) = \begin{cases} 0 \ c_n + \sigma_0^2 \left[ 1 - \exp \left( -\frac{</td>
</tr>
<tr>
<td></td>
<td>$\gamma_e(h) = \begin{cases} 0 \ c_n + \sigma_0^2 \left[ 1 - \exp \left( -\frac{</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$\gamma_e(h) = \begin{cases} 7 \left( \frac{h}{a_0} \right)^2 - 35 \left( \frac{h}{a_0} \right)^3 + \frac{3}{4} \left( \frac{h}{a_0} \right)^5 - \frac{3}{32} \left( \frac{h}{a_0} \right)^7 \end{cases}$, for $h \leq a_0$</td>
</tr>
<tr>
<td></td>
<td>$\gamma_e(h) = \begin{cases} 0 \ c_n + \sigma_0^2 \left[ \frac{3}{2} \frac{</td>
</tr>
<tr>
<td></td>
<td>$\gamma_e(h) = \begin{cases} 0 \ c_n + \sigma_0^2 \left[ \frac{15}{8} \frac{h}{a_0} - \frac{3}{4} \left( \frac{h}{a_0} \right)^3 + \frac{3}{8} \left( \frac{h}{a_0} \right)^5 \end{cases}$, for $h \leq a_0$</td>
</tr>
<tr>
<td></td>
<td>$\gamma_e(h) = \begin{cases} 0 \ c_n + \sigma_0^2 \left[ \frac{3}{4} \left( \frac{h}{a_0} \right)^3 + \frac{3}{8} \left( \frac{h}{a_0} \right)^5 \end{cases}$, for $h &gt; a_0$</td>
</tr>
</tbody>
</table>
Figure F-2. An example of exponential model fitting, with model ideal value, samples (binned), and average.
Appendix G: Spherical harmonic (SPHARM) representation for 3D objects

The spherical harmonic (SPHARM) description is a hierarchical, global, multi-scale boundary description that can only represent objects of spherical topology (Brechbühler et al., 1995). The basic functions of the parameterized surface are spherical harmonics (Gerig et al., 2001). SPHARM can be used to express shape deformations. Truncating the spherical harmonic series at different degrees results in object representations at different levels of detail. SPHARM is a smooth, accurate fine-scale shape representation, given a sufficiently small approximation error.

SPHARM is worked on under a spherical topological assumption. In other word, given a 3D objects, with continuous deformations including stretching and bending, but not tearing or gluing, this 3D object can be deformed as a unit sphere. As illustrate in Figure G-1, (a) is two objects topologically equivalent to a sphere, while (b) is two another two objects topologically equivalent to a torus. Thus, SPHARM can be applied to (a) but not (b).

Mathematically, the SPHARM is similar to the 1D Fourier frequency decomposition, which expresses a random shape wave as a sum of cosine and sine decomposed wavelets. SPHARM, with similar idea, decomposes a 3D object into a sum of spherical harmonics. These spherical harmonics are functions on the sphere that play an analogous role to the cosine and sine functions on a 1D wave. As illustrated in Figure G-2, a 3D object on left can be decomposed as a sum of a constant component, a 1st order component, a 2nd order component, and more.

Based on Figure G-2, spherical harmonic basis functions $Y_{l}^{m}$, $-l \leq m \leq l$ of degree $l$ and order $m$ are defined on $\theta \in [0; \pi] \times \phi \in [0; 2\pi)$ by the following definitions (Brechbühler et al., 1995):

\[
Y_{l}^{m}(\theta, \phi) = \sqrt{\frac{2l + 1}{4\pi} \frac{(l - m)!}{(l + m)!}} P_{l}^{m}(\cos \theta) e^{im\phi}
\]  

(A-5)

\[
Y_{l}^{-m}(\theta, \phi) = (-1)^{m} Y_{l}^{m}(\theta, \phi)
\]  

(A-6)
Figure G-1. Two different types of topology: (a) sphere, (b) torus.
Figure G-2. Harmonic Representation. A 3D object is decomposed as a series of rotate invariant components.
where \( Y_l^m \) denotes the complex conjugate of \( Y_l^m \) and \( P_l^m \) describes the associated Legendre polynomials:

\[
P_l^m(w) = \frac{(-1)^m}{2^l l!} (1 - w^2)^{\frac{m}{2}} \frac{d^{m+l}}{dw^{m+l}} (w^2 - 1)^l
\]

where \( w = \cos \theta \).

To express a surface using spherical harmonics, the three coordinate functions are decomposed and the surface \( v(\theta, \phi) = (x(\theta, \phi), y(\theta, \phi), z(\theta, \phi))^T \) takes the form:

\[
v(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_l^m Y_l^m(\theta, \phi)
\]

where the coefficients \( c_l^m \) are three-dimensional vectors due to the three coordinate functions. The coefficients \( c_l^m \) are obtained by solving a least-squares problem.

The surface expressed as \( v(\theta, \phi) \) usually employed a spherical parameterization, where random shape of 3D objects topologically equivalent to a sphere is deformed into a unit sphere. Different methods have been used to deform a random 3D shape into a sphere. For example, a physical analogy is heat conduction (Brechbühl et al., 1995), which heat the south pole up to temperature \( \pi \), cool the north pole to temperature 0 and ask for the stationary temperature distribution on the heat-conducting surface. As usual in the discrete case, the Laplacian is approximated by finite second differences of the available direct neighbours, which in this sample case implies that every node's latitude (except the poles') must equal the average of its neighbours' latitudes.

Another strategy to mapping a random 3D surface onto a unit sphere is also heat diffusion. But the sphere is larger than the 3D object and it makes sure the 3D object is inside the sphere. With the sphere as heat sink and the 3D object as heat source, isotropic heat diffusion (Chung et al., 2007) is performed for long time. After sufficient amount of time, we reach a steady state, which is equivalent to solving the Laplace equation. A detailed example is given in Figure G-4.
Figure G-3. The processing of “heat diffusion”. Adapted from Brechbühler et al. (1995): (a) Latitude is mapped on the object's surface as a grey value in a; iso-latitude lines are drawn every $\pi /16$. (b) Longitude is shown, iso-longitude lines (“meridians”) are $\pi /8$ apart.
Figure G-4. The process of using isotropic heat diffusion to map a 3D object onto a larger sphere. Left: The 3D object in the centre is assigned the value 1 and the sphere enclosing the object is assigned the value -1. Right: After solving isotropic heat diffusion for long time, we reach a steady state, which can be used to generate a mapping from the amygdala surface to the sphere by taking the geodesic path from value 1 to -1.
Appendix H: Matlab codes for 3D objects comparison

function [Comp,Corr,Q, rmsd,dxdydz,distall]=Building3DEvaluation(refSTLfile, smpSTLfile)
% -----------------3D building model evaluation -------------------------
% Based on:
% Chuiqing Zeng, et al.,"An evaluation system for 3D building reconstruction".
% Project description:
% the shape similarity of 3D objects is comparison between a sample and
% a reference object using a distance measure.
% In this study, 3D buildings are the objects for the comparison.
% Reference 3D buildings comes from ground survey or other maps, while the
% extracted buildings come from remote sensing image reconstruction algorithms
% The purpose of this code is to measure the accuracy, or how similar the
% reconstructed building is when compared with the ground-surveyed ones.

% Algorithm Description:
% For a reference 3D building (ref) and a detected building sample (smp),
% the similarity is divided into three components: the 3D volume difference
% , the 3D shape comparison based on the surface parameterization, and
% the 3D corner/feature points difference. Each component describes
% the accuracy from a different perspective.

% Metrics for evaluation
% (1)For the Volume difference, random points will be used to measure
% traditional User/Producer Accuracy, as well as Quality.
% (2)For the shape similarity, SPHARM coefficients are used to describe buildings
% for the comparison, more details can be found: http://pages.stat.wisc.edu/~mchung/research/amygdala/
% (3)For 3D feature point distance: Euclidean distance in 3D space is used to
% measure the corner/feature point shift. RMSD, MEAN, STD are used for
% evaluation.

% Input and Output:
% Input: the two building models in STL 3D file format. More details
% about STL can be found: http://en.wikipedia.org/wiki/STL_(file_format)
% STL is a format without surface color supporting, which fits our research
% , where no color information is concerned in detected buildings yet.
% Output: the metrics for the three components to evaluate the accuracy.

% Syntax:
% [Comp,Corr,Q, rmsd,dxdydz,dist] = StartComparison (refSTLname ,
% smpSTLname)
% Arguments:
% refSTLname , smpSTLname -input STL file names for reference and
% sample buildings
% Returns: Comp, Corr, Q - random sampled evaluated accuracy
%          rmsd - metrics for shape similarity after 3D
%          sphere parameterization and SPHARM coeff comparison.
%          dx, dy, dz, dist - shift on x, y, z direction and distance between the
%          two compared building models.

% Examples:
%     [Comp, Corr, Q, rmsd, dx, dy, dz, dist] = StartComparison('reference.stl', 'sample.stl')
%
% See also: LoadSTL, PointInTriangles, Project_Comp, Distance_cornerPTs

% Copyright (c) 2013 Chuiqing Zeng
% University of Western Ontario, Canada

% Permission is hereby granted, free of charge, to any person obtaining a copy
% of this software and associated documentation files (the "Software"), to deal
% in the Software without restriction, subject to the following conditions:

% The above copyright notice and this permission notice shall be included in
% all copies or substantial portions of the Software.
% The Software is provided "as is", without warranty of any kind.

% Load stl data for both reference and sample buildings
[refFV, refnorm, meanX, meanY] = LoadSTL(refSTLfile);
smpFV = LoadSTL(smpSTLfile, meanX, meanY);

% STEP01: 3D volume difference by random points inside ref/sample buildings
PtNum = 2000;
% test whether a point is inside a 3D object
[PTin, PTlist] = PointInTriangles(refFV, smpFV, PtNum);
total = PTin(:, 1) + PTin(:, 2);
Comp = sum(total == 2) / sum(PTin(:, 1) > 0)
Corr = sum(total == 2) / sum(PTin(:, 2) > 0)
Q = sum(total == 2) / sum(total > 0)

% STEP02: 3D shape similarity between two buildings in this step will be implemented separately.
% dirName = The-path-to-the-result-assigned-by-function-Weighted_SPHARM
rmsd = Building3Dshapesimilarity(dirName);

% STEP03: 3D feature/corner point distance by calculating distance for corner points.
[refPT, smpPT, diff, dist] = Distance_cornerPTs(refFV, smpFV, 0.98);
% compute the Mean and std for DeltaX, Y, Z
diff = diff';
dx, dy, dz = sqrt(sum(diff.^2) / size(diff, 1))
distall = mean(dist);

function [PTin, PTlist] = PointInTriangles(refFV, smpFV, PtNum)
% ----- This function is a part of the 3D building evaluation process ------

function [PTin,PTlist]= PointInTriangles(refFV,smpFV, PtNum)
% -----This function is a part of the 3D building evaluation process-------
% Description:
% This function is used to Test whether a point is inside a 3d object.
% 
% Usage:
% [PTin,PTlist]= PointInTriangles(refFV,smpFV, PtNum)
% 
% Arguments:
% refFV ---structure--the face and vertices structure for the
% reference building
% smpFV ---structure--the face and vertices structure for the
% sample building
% PtNum ---scalar ---the number of points as random samples
% Returns: PTin ---N*2--the output result for the points whether they are
% inside the 3D object. N is the size of random sample (PtNum here).
% PTlist ----N*3 ----the 3D random point list been used.
% 
% Examples:
% PTin=PointInTriangles(refFV,smpFV, 1000)
% See also: StartComparison,LoadSTL,project_comp,Distance_cornerPTs

% given a search box and define random points in the box
minCoords=min([refFV.vertices;smpFV.vertices]);
maxCoords=max([refFV.vertices;smpFV.vertices]);
box=[minCoords;maxCoords];

%generate random points
Gain=ones(PtNum,1)*(maxCoords-minCoords);
Offset=ones(PtNum,1)* minCoords;
PTlist=rand(PtNum,3).*Gain + Offset;

refvert1= refFV.vertices(refFV.faces(:,1),:);
refvert2= refFV.vertices(refFV.faces(:,2),:);
refvert3= refFV.vertices(refFV.faces(:,3),:);
smpvert1= smpFV.vertices(smpFV.faces(:,1),:);
smpvert2= smpFV.vertices(smpFV.faces(:,2),:);
smpvert3= smpFV.vertices(smpFV.faces(:,3),:);

% test the point one by one to see whether it is inside the triangles
PTin=zeros(PtNum,2);
for idx =1: PtNum
    PTin(idx,1)= IsInVolume(PTlist(idx,:),refvert1,refvert2, refvert3);
    PTin(idx,2)= IsInVolume(PTlist(idx,:),smpvert1,smpvert2, smpvert3);
end
end
function in = IsInVolume(pt,vert1,vert2, vert3)
% -----This function is a part of the 3D building evaluation process-------
% 
% Description:
% the strategy here used is given a point and test whether there is 
% intersection with the STL model in 6 directions (+/-X, +/-Y, and +/-Z) 
% only when ALL directions have intersected surfaces, then the point is 
% labeled as inside the volume 
% 
% Usage:
% in = IsInVolume(pt,vert1,vert2, vert3)
% 
% Arguments:
% pt ---3*1---an point for test 
% vert1,2,3 ---N*3--point list for the triangles constructed the 3D 
% object. each triangle with three points vert1,vert2, vert3. 
% 
% Returns:
% in ---N*1--a Boolean list for the point test result, 
% with 0 represents not in and 1 means in the object. 
% 
% Examples:
% in = IsInVolume([0 0 0],vert1,vert2, vert3)
% See also: StartComparison,LoadSTL,project_comp,Distance_cornerPTs

theta=[0 30 60 120 150];
dx=cos(pi/180.*theta);
dy=sin(pi/180.*theta);
dirXY=[dx;dy;zeros(1,5)]';
dirXZ=[dx;zeros(1,5);dy]';
dirYZ=[zeros(1,5);dx;dy]';
dir=[dirXY;dirXZ;dirYZ];
in=1;
for i=1:length(dir)
    if IsIntersected(pt,dir(i,:), vert1,vert2, vert3)==0
        in=0;
        break;
    end
end

end

function Num= IsIntersected(pt, dir, vert1,vert2, vert3)
%test whether a ray is intersect with volume

Orig  = repmat(pt,size(vert1,1),1); % Clone it until the same size as vert1
Dir  = repmat(dir,size(vert1,1),1); % Clone it until the same size as vert1
% the core intersection function.
% this function can be downloaded from Matlab resource centre, provided by Jarek Tuszynski (jaroslaw.w.tuszynski@saic.com)
intersect = TriangleRayIntersection(Orig, Dir, vert1, vert2, vert3); %, option
Num = sum(intersect);
end

function [surf_smooth, fourier] = Weighted_SPHARM(strVTKpath, nIterator, degree, Sigma)
% e.g.: [surf_smooth, fourier] = Weighted_SPHARM('test.vtk', 10, 80, 1e-3)
% the following codes are used to load a building as voxel data, and compute corresponding RMSD of SPHARM coefficients for comparison----------------
[bim, origin, vxsize] = readvtk(strVTKpath);
vol = logical(bim);

% Surface parameterization: mapping a random 3D object onto a unit sphere
% more details can found: http://pages.stat.wisc.edu/~mchung/research/amygdala/
[amyg, sphere, amygisphere] = CREATEenclosedamyg(vol, surf);

% after enough time, it reaches a steady state, which is equivalent to solving the Laplace equation.
% after 5 times iteration in this case:
stream = LAPLACE3Dsmooth(amygsphere, amyg, -sphere, nIterator, 5, 10);

% draw the pass of "heating": from 1 --> -1
sphere = isosurface(amyg);
for alpha = 1:nIterator
    sphere = LAPLACEcontour(stream, sphere, 1 - 2*alpha/nIterator);
    sphere = REGULARIZEarea(sphere, 0.8);
end;

surf = isosurface(amyg);
% weighted spherical harmonic representation of degree 60 and bandwidth sigma=0 is given by running the code
[surf_smooth, fourier] = SPHARMsmooth2(surf, sphere, degree, Sigma);

% save the result for further comparison between reference and sample buildings
save(strrep(strVTKpath, '.vtk', '.mat'))
end

function rmsd = Building3Dshapesimilarity(dirName)
% the following scripts are used to conduct 3D building shape similarity evaluation based on the comparison between reference and sample building coefficients
% dirName='D:\Matlab\Work\SpharmMatDir\Comparison_stl';
files = dir(fullfile(dirName, 'ref_*.mat'));
files = {files.name...
degree = 20;
rmse = zeros(numel(files), 1);
for i = 1:numel(files)

    % read the reference file and its SPHARM coeffs
    refName = fullfile(dirName, files{i});  %# full path to file
    load(refName)
    fourier_ref = SPHARMvectorize_single(fourier, degree);

    % read the sample file and its SPHARM coeffs
    smpName = strrep(refName, 'ref', 'smp');
    load(smpName);
    fourier_smp = SPHARMvectorize_single(fourier, degree);

    % calculate the norm distances based on SPHARM coeffs
    dist = [fourier_ref.x - fourier_smp.x; fourier_ref.y - fourier_smp.y; fourier_ref.z - fourier_smp.z];
    temp = norm(dist(:))/sqrt(4*pi)
    rmse(i) = temp;
end
end

function [refPT, smpPT, diff, dist] = Distance_cornerPTs(refFV, smpFV, T_norm, isDraw)
% -----This function is a part of the 3D building evaluation process-------
% Description:
% calculate the distance between two point sets.
% the refPT are point sets from reference building corners, while
% the smpPT are point sets from sample building corners
% output: the difference vector and distance
%
% Usage:
% [outXY,OutlinePts,P]=TrianglesUnion(FV, dir, draw)
%
% Arguments:
% refFV ---structure--the face and vertices structure for the
% reference building
% smpFV ---structure--the face and vertices structure for the
% sample building
% T_norm ---scalar ---the threshold for the norm of final norm
% after CROSS operation for the three faces (F1-3),
% T_norm=asin(cross(norm(cross(norm_F1,norm_F2)),norm_F3))
% isDraw ---anything ---to draw the result or not, can be any
% number or text.
% Returns:
% refPT ---N*3--a list of point to represent in the reference
% building after filtering. only CORNER pts are kept.
% smpPT ---N*3---a list of point to represent in the sample
% building after filtering. Only CORNER pts are kept.
% diff ----N*3----the difference vector between refPT and smpPT
% Z for the last dimension.
% dist ----N*1----Euclidean distance (1D norm) of diff.

% Examples:
% [refPT, smpPT,diff,dist]=Distance_cornerPTs(refFV, smpFV,0.95,'plot')
% See also: StartComparison, LoadSTL, PointInTriangles, Project_Comp

if nargin == 3  % default not draw and show the results
    isDraw=0;
elseif nargin == 2
    T_norm=0.1;
    isDraw=0;
else
    isDraw=1;
end

% Simplify the triangles and search for the corner points
refPT=SearchCornerPTs(refFV,T_norm);
smpPT=SearchCornerPTs(smpFV,T_norm);
refPT=refPT';
smpPT=smpPT';

% The comparison function, search for the nearest points
% This function can be downloaded from Matlab resource centre,
% Copyright 2006 Richard Brown
IdxMatched = nearestneighbour(refPT, smpPT);
smpPT=smpPT(:,IdxMatched);
diff=smpPT -refPT;
dist=sqrt(diff(1,:).^2+diff(2,:).^2+diff(3,:).^2);

% Sort all the correspond points according to their distance
[dist,idx]=sort(dist);
refPT=refPT(:,idx);
smpPT=smpPT(:,idx);
diff=diff(:,idx);

% Find unique point matches and then keep the shortest corresponds.
[uniquesmp,uniqueList] =unique(IdxMatched(idx),'first');
refPT=refPT(:,uniqueList);
smpPT=smpPT(:,uniqueList);
diff=diff(:,uniqueList);
dist=dist(:,uniqueList);

% Draw the figures and results.
if isDraw  % draw the figure
    figure;
    patch(smpFV, 'FaceColor',[0.9,0.9,0.9], 'EdgeColor', 'none'), camlight
    axis equal
hold on
plot3(smpPT(1,:), smpPT(2,:),smpPT(3,:), 'b^','MarkerSize', 5,'MarkerFaceColor','b');
plot3(refPT(1,:), refPT(2,:),refPT(3,:), 'r.', 'MarkerSize', 15,'LineWidth',1);
quiver3(refPT(1,:), refPT(2,:),refPT(3,:), diff(1,:), diff(2,:),diff(3,:), 0,'k','LineWidth',1);
hold off
xlabel('East','FontSize',16,'FontName','Times New Roman');
ylabel('North','FontSize',16,'FontName','Times New Roman');
zlabel('Height','FontSize',16,'FontName','Times New Roman');
legend('3D building','sample','reference','diff vector')
end
end

function [crnPT,CornerPTList]=SearchCornerPTs(FV,T_norm)
% ---------------------function ------------------------
% search the corner point from the list of vertices in a FV structure.
% %
% % Goal:
% % To search for the corner points in the face/vertices structure.
% %
% % Description:
% % input the FV structure and output the corner points.
% % the idea: a corner should be at the intersect of three planes which is
% % almost perpendicular to each other.
% %
% % Arguments:
% % FV ---structure--the face and vertices structure for the
% % reference/sample building
% % T_norm ---scalar ---the threshold for the norm of final norm
% % after CROSS operation for the three faces (F1-3).
% %
% Returns:
% % crnPT ---N*1--a list of point index from the FV to represent in the reference
% % building after filtering. only CORNER pts are kept.
% % CornerPTList ---the true coordinates for the crnPT.
% %
% Examples:
% [crnPT,CornerPTList]=Distance_cornerPTs(refFV,0.95)

if nargin < 2  %default not draw
    T_norm=0.95;
end;

Vertices=FV.vertices;
Faces=FV.faces;

% compute the norm
u=cross(Vertices(Faces(:,2,:))-Vertices(Faces(:,1,:)),Vertices(Faces(:,3,:))-Vertices(Faces(:,1,:)));
temp = sqrt(u(:,1).^2+u(:,2).^2+u(:,3).^2);
u = u./repmat(temp,1,3); % Make u of unit length (normalization)

% convert the N*3 facets of vertices into a long list to search for each vertex
temp = reshape(Faces',[],1);
val = unique(temp);

CornerPTList = zeros(length(Faces),1);
% circulate and determine each vertices to see if it is a corner pt
for idx = 1:length(val)
    blsCorner = 0;
    faceIDX = ceil(find(temp==val(idx))/3);
    % if there is no enough faces at these vertices, then skip
    if length(faceIDX) <= 3
        continue;
    end
    % otherwise will check whether the faces are perpendicular
    for i = 1:length(faceIDX)-2
        for j = i+1:length(faceIDX)-1
            tempNorm = cross(u(faceIDX(i,:),:),u(faceIDX(j,:),:));
            if norm(tempNorm) > (T_norm/10) % ignore almost parallel vectors
                tempNorm = tempNorm./norm(tempNorm);
                for k = j:length(faceIDX) % min([j+1,length(faceIDX)])
                    if norm(cross(tempNorm,u(faceIDX(k,:)))) < T_norm
                        % it is a valid corner point
                        blsCorner = 1;
                        break;
                    end
                end
            end
        end
    end
    if blsCorner
        break;
    end
end
% determine whether the point need to add to the list
CornerPTList(idx) = blsCorner;
end
CornerPTList = val(CornerPTList>0);
crnPT = Vertices(CornerPTList,:);
end
Chapter 3:


From: **Canadian Journal of Remote Sensing Editor in Chief** <editor.cjrs@hotmail.com>
Date: Sun, Mar 9, 2014 at 8:27 PM
Subject: RE: Canadian Journal of Remote Sensing - Decision on Manuscript ID CJRS-12-0111.R1
To: "C. Zeng"

Dear Chuiqing

With this email I grant you permission to use the published paper in Canadian Journal of Remote Sensing in your thesis.

Yours sincerely,

Nicholas Coops

Canada Research Chair in Remote Sensing

Chapter 2 and Chapter 4:


The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you must follow the requirements listed below:

Full-Text Article

If you are using the entire IEEE copyright owned article, the following IEEE copyright/credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]

Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.

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

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

References for Appendices


CV: Chuiqing Zeng

EDUCATION

PhD Candidate, Geography
The University of Western Ontario, London, ON, Canada
Sept. 2010-May 2014

Master of Science, Cartography and Geographic Information System
Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China
Sept 2007-July 2010

Bachelor of Engineering, Remote Sensing Science and Technology
Wuhan University, Wuhan, Hubei, China
Sept 2003-July 2007

PUBLICATIONS

First-author peer reviewed papers:


Conference papers:


CONFERENCE PRESENTATIONS


