Universität der Bundeswehr München

Institut für Angewandte Informatik

From LIDAR Point Clouds to 3D Building Models

Dissertation

Hossein Arefi

2009

Universität der Bundeswehr München
Fakultät für Bauingenieur- und Vermessungswesen

From LIDAR Point Clouds to 3D Building Models

M.Sc. Hossein Arefi

Vollständiger Abdruck der von der Fakultät für Bauingenieur- und Vermessungswesen der Universität der Bundeswehr zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Promotionsausschuss:

Vorsitzender: Univ.-Prof. Dr.-Ing. Friedrch Kröll
1. Berichterstatter: Univ.-Prof. Dr.-Ing. Helmut Mayer
2. Berichterstatter: Univ.-Prof. Dr.-Ing. Uwe Stilla, Technische Universität München
3. Berichterstatter: PD Dr.-Ing. habil. Johannes Engels, Universität Stuttgart

Diese Dissertation wurde am 15. Mai 2009 bei der Universität der Bundeswehr München eingereicht.
Tag der mündlichen Prüfung: 25. September 2009

Acknowledgments

I would like to express my utmost gratitude to my major advisor Prof. Dr.-Ing. Helmut Mayer for his professional guidance, direction, and encouragement during my thesis. Without his assistance this dissertation would have never been accomplished.

I am also grateful to Prof. Dr. Uwe Stilla and Dr. Johannes Engels for their helpful advice as my committee members. Their direction and guidance, especially on improving the quality of my dissertation, were very useful.

I am also highly thankful to Prof. Dr. Friedrch Kröll, chair of the PhD committee, for his valuable suggestions.

I would like to express my profound gratitude to Prof. Dr. Michael Hahn, for his invaluable support, encouragement, and useful suggestions during my research work at Stuttgart University of Applied Sciences. His moral support and continuous guidance enabled me to complete my work successfully. I thank my colleagues at University of Applied Sciences for scientific, financial, and administrative support.

I would especially like to thank Dr. Peter Reinartz, Head of the Department of Photogrammetry and Image Analysis at German Aerospace Center (DLR Oberpfaffenhofen), for his guidance and for giving me time to complete my thesis. I thank colleagues at the DLR for their comprehensive help, particularly in the last year of my research.

I am grateful for the financial support provided by the Federal Ministry of Education and Research (BMBF) as well as by Dr. Joachin Lindenberger (TopScan GmbH) during my research and study at Stuttgart University of Applied Sciences.

Last but not least, I would like to thank God for blessing me with the health and strength to complete this process and for surrounding me with smart and dedicated people who have supported me all the way, especially my parents and my kind family. My wife and children have always been a source of joy and comfort throughout my PhD study. I want to thank them for all the support they gave me and ask them to forgive me all the time I did not spend with them.

Dedication

To my beloved family

Tahereh, Mohammad Mahdi & Reihaneh

Abstract

High quality and dense sampling are two major properties of recent airborne LIDAR data which are still improving. In this thesis a novel approach for generating 3D building models from LIDAR data is presented. It consists of four major parts: filtering of non-ground regions, segmentation and classification, building outline approximation, and 3D modeling.

With filtering non-ground structures are eliminated from the laser data. Image reconstruction by means of geodesic morphology is at the core of the proposed algorithm. Structures which do not comply concerning size or shape are suppressed. By interpolating the bald earth produced by filtering, Digital Terrain Models (DTM) are generated. Image segmentation creates potential non-ground regions which are subject to rule-based classification. Geometric feature descriptors based on surface normals, the local height variation, and a vegetation index are employed to classify data into buildings, trees, and other objects such as power lines and cranes.

After building classification, their outlines are extracted and unnecessary points are eliminated by two approximation procedures. One fits rectilinear polygons to the building outlines by a hierarchical adaptation of Minimum Bounding Rectangles (MBR). This works fast and reliably, but is restricted to rectangular shapes. For non-rectangular polygons, a Random Sample Consensus (RANSAC) based procedure is employed to fit straight lines. Lines are then intersected or joined.

The automatic generation of 3D building models follows the definitions of the Levels of Detail (LOD) in the CityGML standard. Three LOD are considered in this thesis. The first LOD (LOD0) consists of the extracted DTM from the LIDAR data. A prismatic model containing the major walls of the building forms the LOD1. For it, the building roof is approximated by a horizontal plane. LOD2 includes the roof structures into the model. A model driven approach based on the analysis of the 3D points in 2D projection planes is proposed to analyze the roof structure. Building regions are divided into smaller parts according to the direction and the number of ridge lines, the latter extracted using geodesic morphology. A 3D model is derived for each building part. Finally, a complete building model is formed by merging the 3D models of the building parts and adjusting the nodes after merging.

Results for test data show the potential, but also the shortcomings of the approach also in comparison to related work.

Zusammenfassung

Hohe Genauigkeit und Punktdichte sind zwei bedeutende Eigenschaften von derzeitigen luftgestützten LIDAR-Daten, die immer besser werden. In der vorliegenden Arbeit wird ein neuartiger Ansatz zur Erzeugung von 3D-Gebäudemodellen aus LIDAR-Daten vorgestellt. Er besteht im Wesentlichen aus vier Teilen: Filterung von Flächen, die nicht auf dem Gelände liegen, Segmentierung und Klassifizierung, Approximation von Gebäudeumrissen und 3D-Modellierung.

Durch Filterung werden nicht auf dem Gelände liegende Strukturen aus den Laserdaten entfernt. Die Bildrekonstruktion mittels der geodätischen Morphologie bildet den Kern des vorgeschlagenen Algorithmus. Strukturen, die nicht gegebenen Anforderungen bezüglich Größe oder Form genügen, werden unterdrückt. Durch Interpolation der mittels Filterung bestimmten auf dem Gelände liegenden Punkte werden Digitale Geländemodelle (DGM) generiert. Bild-Segmentierung erzeugt potenzielle nicht auf dem Gelände liegenden Punkte, die einer regelbasierten Klassifizierung unterworfen werden. Deskriptoren für geometrische Merkmale basierend auf Oberflächennormalen, der lokalen Höhenvariation sowie einem Vegetationsindex werden verwendet, um die Bildsegmente als Gebäude, Bäume, oder andere Objekte, wie z.B. Stromleitungen oder Kräne, zu klassifizieren.

Nach der Klassifikation der Gebäude werden ihre Umrisse extrahiert und unnötige Punkte mit Hilfe zweier Approximationsprozeduren eliminiert. Eine passt mittels einer hierarchischen Adaption von minimalen umschließenden Rechtecken (MBR) rechtwinklige Polygone an die Gebäudeumrisse an. Dies funktioniert schnell und zuverlässig, ist jedoch auf rechtwinklige Grundrisse beschränkt. Für nicht-rechtwinklige Polygone wird eine "Random Sample Consensus" (RANSAC) basierte Prozedur benutzt, um Geradenstücke anzupassen. Diese Geradenstücke werden anschließend verschnitten oder verbunden.

Die automatische Erzeugung von 3D-Gebäudemodellen folgt den Definitionen der "Levels of Detail" (LOD) im CityGML Standard. Drei LOD, also Detaillierungsstufen, werden in dieser Arbeit verwendet. Der erste LOD (LOD0) besteht aus dem DGM, das aus den LIDAR-Daten extrahiert wird. Ein prismatisches Modell, welches die wichtigsten Wände des Gebäudes enthält, bildet den LOD1. Für dieses wird das Gebäudedach durch ein horizontales Ebenenstück approximiert. LOD2 nimmt die Dachstruktur in das Modell auf. Es wird ein modellgetriebener Ansatz zur Analyse der Dachstruktur vorgeschlagen, der auf der Analyse der 3D Punkte in 2D Projektionsebenen basiert. Gebäuderegionen werden entsprechend Richtung und Zahl der Firstlinien in kleinere Bereiche unterteilt, wobei die Firstlinien mit Hilfe der geodätischen Morphologie extrahiert werden. Für jedes Gebäudeteil wird ein 3D-Modell abgeleitet. Schließlich wird ein komplettes Gebäudemodell durch Verschmelzung der 3D-Modelle der einzelnen Gebäudeteile und einer anschließenden Ausgleichung der Knoten gebildet.

Ergebnisse für Testdaten zeigen das Potential, aber auch die Schwächen des Ansatzes auch im Vergleich zu verwandten Arbeiten.

Contents

1 Introduction **9**

1.1 Motivation . . . 9

1.2 Research Focus and Methodology . . . 9

1.3 Structure of the Thesis . . . 11

2 Basic Terms and Techniques **13**

2.1 Airborne Laser Scanning . . . 13

2.2 Range Images and Range Features . . . 15

2.2.1 Gradient Image . . . 16

2.2.2 Surface Normals . . . 17

2.2.3 Gaussian and Mean Curvature . . . 18

2.3 Morphological Reconstruction Using Geodesic Dilation . . . 18

2.3.1 Basic Operations of Gray-scale Mathematical Morphology . . . 18

2.3.2 Gray-scale Reconstruction Based on Geodesic Dilation . . . 21

2.4 Robust Model Fitting . . . 26

2.4.1 Hough Transform . . . 27

2.4.2 Random Sample Consensus . . . 28

2.5 Building Models . . . 30

2.5.1 Roof Types and Modeling of Buildings . . . 30

2.5.2 Levels of Detail . . . 32

3 Former Work **34**

3.1 3D Building Reconstruction from LIDAR Data . . . 34

3.1.1 Filtering of Non-ground Regions and DTM Generation . . . 34

3.1.2 Building Outline Detection and Approximation . . . 36

3.1.3 3D Building Models . . . 38

3.2 3D Building Reconstruction Using Other Data Sources . . . 39

3.2.1 3D Building Reconstruction from Imagery . . . 39

3.2.2 3D Building Detection and Reconstruction by Combination of LIDAR and Other Data Sources . . . 42

4 A Novel Approach for Filtering of LIDAR Range Data Based on Geodesic Morphology 43

4.1 Pre-processing . . . 43

4.1.1 Outlier Detection and Removal . . . 43

4.1.2 Data Rasterization . . . 46

4.2 Separation of Ground and Non-ground Based on Morphological Reconstruction 50

4.3 Generation of Normalized DSM and DTM . . . 56

5 Detection and Approximation of Building Outlines 60

5.1 LIDAR Data Classification . . . 61

5.1.1 Extraction of Feature Descriptors . . . 61

5.1.2 Segmentation of Non-ground Regions . . . 63

5.1.3 Rule-based Classification . . . 63

5.2 Building Outline Approximation . . . 68

5.2.1 Boundary Approximation by Hierarchical Fitting of Rectangular Polygons 69

5.2.2 Boundary Approximation Using RANSAC . . . 73

5.2.3 Comparison of Methods for Building Outline Approximation . . . 75

6 Projection-Based 3D Building Modeling 79

6.1 Automatic DTM Generation – LOD0 . . . 79

6.2 Building Outline Detection and Approximation for Generating 3D Prismatic Models – LOD1 . . . 81

6.3 A Novel Approach for Building Reconstruction Using Projection Based Analysis of 3D Points – LOD2 . . . 82

6.3.1 Ridge Line Extraction and Main Orientation Determination . . . 84

6.3.2 Localization of Building Parts . . . 85

6.3.3 Projection from 3D into 2D and 2D Model Fit . . . 85

6.3.4 From 2D to 3D – LOD2 . . . 87

7 Experimental Investigations with Stuttgart Test Data 91

7.1 Test Data . . . 91

7.2 Experimental Results . . . 91

7.2.1 Separation of Non-ground and Ground Regions – Generation of DTM and nDSM . . . 93

7.2.2 Classification of LIDAR Data and Approximation of Building Outlines . 102

7.2.3 3D Building Modeling on LOD2 . . . 107

7.3 Discussion of Results . . . 114

7.3.1 Generation of LOD0 . . . 114

7.3.2 Generation of LOD1 . . . 114

7.3.3 Generation of LOD2 . . . 115

8 Conclusions and Recommendations **117**

8.1 Summary and Conclusions 117

8.2 Recommendations 119

Chapter 1

Introduction

1.1 Motivation

Recently, three dimensional (3D) digital representations of real world objects become increasingly important. Particularly, 3D city models can be employed for a variety of applications such as telecommunication, urban planning, environmental simulation, cartography, tourism, and mobile navigation systems. They consist of geometric 3D models for buildings, roads, trees, etc. and possibly high resolution aerial or satellite image data. Therefore, researchers in the fields of geo-informatics, especially in Photogrammetry and Geographic Information Systems (GIS) are interested in building extraction and 3D reconstruction. Building reconstruction based on aerial imagery focuses among other issues on generating Digital Surface Models (DSM) by image matching used to extract the geometric properties of buildings.
Buildings come with a diversity of shapes and details. Depending on the existing data quality, e.g., point resolution, the 3D model can be produced with different levels of detail (LOD). The roof of the building is almost the only part of a building which is visible in airborne photogrammetric and remote sensing data. Therefore, the algorithms for the 3D modeling of buildings start with the extraction of the roofs' geometry and then extend them by integrating the walls to form complete 3D building models.
Since the 1990*s* LIDAR data are employed for building extraction due to the high quality of the automatically generated DSM. High accuracy as well as high density of the laser points are the main reasons why LIDAR data are often preferred to DSM generated by photogrammetry. Due to the complexity of the structure of many buildings fully automatic building reconstruction is not feasible yet. Therefore, manual and semi-automated approaches are still used in most present projects.
Because of the high costs of manual or semi-automatic 3D data capture from aerial images, maps, or laser data, which is due to the large number of 3D buildings well as their high degree of complexity, automatic processes appear to be the only viable way to satisfy the needs of the users (FÖRSTNER, 1999). The automatic production of 3D city models is thus one of the major wishes of many clients of Photogrammetry and GIS.

1.2 Research Focus and Methodology

The objective of this thesis is an automatic approach for the creation of 3D building models from LIDAR point clouds. As result, 3D vector Computer Aided Design (CAD) models describing

the structures of building represent the laser points. To achieve this objective, approaches for filtering and classification of the laser points as well as a concept for the modeling of buildings based on the analysis in projection planes are proposed. Figure 1.1 summarizes the overall work

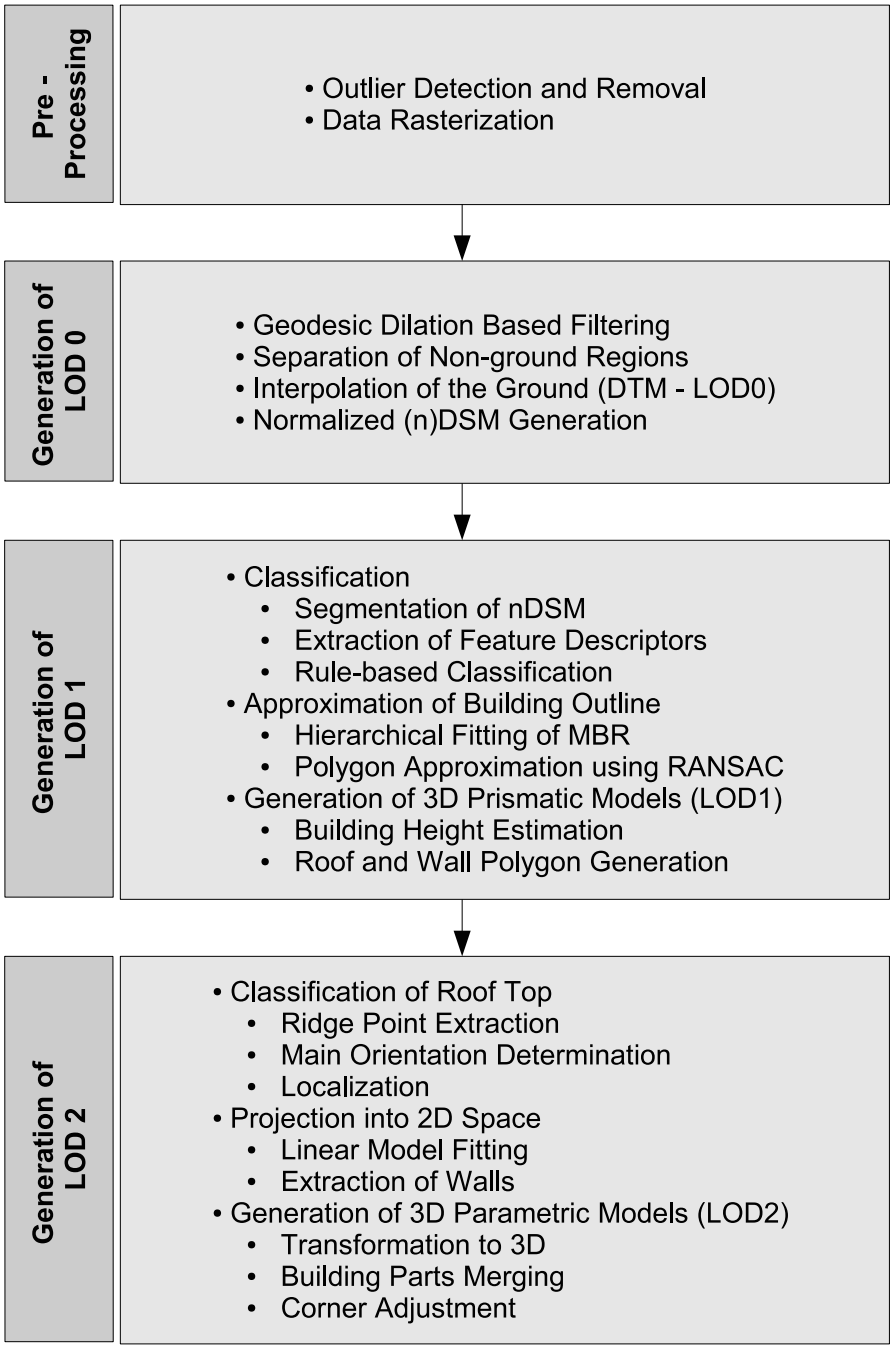


Figure 1.1: Work flow of the proposed algorithm for 3D building model generation from LIDAR point clouds

flow. It starts with pre-processing. The raw laser range data are interpolated after removing outliers to generate a regular matrix representing the range image. In the second phase, the normalized DSM (nDSM) is created via a hierarchical approach. At the core is the extraction of non-ground regions by means of geodesic image reconstruction. Iterative geodesic dilation

is employed to filter the potential non-ground regions. Regions are confirmed as non-ground if they show a significant discontinuity on their boundary. The decision is based on feature descriptors, which can highlight height jumps in the image, such as surface normals or local range variation (LRV). After separating the non-ground regions, the produced gaps are filled by interpolation to shape the Digital Terrain Model (DTM). The nDSM is produced by subtracting the DTM from the original range image.

The third phase consists of segmentation and classification. A threshold-based segmentation is employed to segment the non-ground regions. The latter are rule-based classified into buildings, trees, and other objects such as power lines and cranes. The outlines of buildings are generated using morphological operations. For building outline approximation two methods based on Minimum Bounding Rectangle (MBR) and RANdom SAmple Consensus (RANSAC) are employed for the reduction of the number of points of the building outline. Finally, the novel approach for the automatic 3D modeling of buildings on three different levels of detail is explained. The LOD follow the standard definition of the City Geography Markup Language (CityGML) of (Kolbe et al., 2005). The first LOD (LOD0) corresponds to the DTM. The LOD1 level consists of a 3D representation of buildings using prismatic models based on the approximated building outline. An average height level is estimated from the points located inside the building.

To generate the third level of detail (LOD2), a projection based approach is proposed for reconstructing building models with roof structures. The core idea is to analyze the laser data in projection planes. This phase begins with the classification of the roofs based on the highest points extracted by geodesic dilation, local curvature, LRV, and surface normals. The classification distinguishes roofs with ridge points and flat roofs. The Hough Transform is used to separate the ridge lines if they point in different directions. This step divides complex buildings into building parts with different ridge lines and flat parts. Every individual building part is analyzed separately for 3D model generation. The points belonging to each part are projected to a 2D coordinate system with the axes orthogonal to the ridge direction, the latter calculated in the form of the main orientation of the building part. The points belonging to a building part are defined by a rectangle around the ridge line.

A model is fitted to the data in 2D space. If the building part contains a ridge line, a polyline consisting of five connected lines is used. Two lines represent the vertical walls, another two the gable part of the building, and one line the floor. Similarly, for flat roofs a model is used with four lines, two vertical for the walls and two horizontal for roof and floor. After generating individual models for the building parts, they are merged, and the nodes and corners are adjusted resulting in the final 3D model of the building.

1.3 Structure of the Thesis

The thesis is structured as follows: In Chapter 2 terms and techniques, which are utilized in this thesis, are explained. The chapter begins with a brief introduction into laser scanning systems. The main components of an airborne laser scanner and its properties are reviewed and the different types of data produced with a laser scanner system are explained. The chapter continues by describing geometric feature descriptors which can be derived from LIDAR range images. They are used to characterize different structures in the laser image to improve the separability in image segmentation and classification. An overview of gray-scale mathematical morphology as well as a detailed description of geodesic morphology and image reconstruction based on it is given next, comprising also the unique properties of image reconstruction by

geodesic morphology. The chapter ends with a short overview over two robust approaches for fitting geometric models in the form of 2D and 3D straight lines and/or 3D planes to a point cloud containing outliers.

A brief summary of former work on the processing of imagery and LIDAR data for generation of 3D building models is given in Chapter 3. In Chapter 4 the proposed algorithm for filtering LIDAR range images is outlined. The chapter begins with explaining the pre-processing methods applied to the laser scanning data, particularly the detection and removal of outliers as well as the interpolation of the raster grid. The description continues by explaining the separation of non-ground structures from the ground surface. At the core is the generation of normalized Digital Surface Models (nDSM) by means of a hierarchical approach based on geodesic image reconstruction.

Chapter 5 consists of two major parts: In the first, the proposed method for the classification of the laser image into three classes, namely “building”, “vegetation", and “other" is presented and discussed. After the geometric feature descriptors employed for the distinction of the segmented regions are explained, the rule-based classification, utilizing the geometric descriptors, is discussed. In the second part, the approximation of the building outlines is described. Two techniques are discussed for this purpose, namely a hierarchical approach based on the MBR and a RANSAC-based approach.

Chapter 6 describes a projection-based approach for the 3D modeling of buildings. It begins with the classification of the roofs into “gable roofs" and “flat roofs". The main orientation of each building part is determined and according to the roof type, the corresponding 3D model is fitted in 2D projection space. The building is finalized by merging, refining, and adjusting the roof segments.
In Chapter 7 the experimental investigations with the Stuttgart test data are presented and the results are analyzed. Finally, Chapter 8 summarizes the contributions of the thesis and recommends directions for future research.

Chapter 2

Basic Terms and Techniques

2.1 Airborne Laser Scanning

Airborne laser scanning (ALS), also referred to as airborne LIDAR (Light Detection And Ranging), is a very convenient source of information for extracting Digital Surface Models (DSM). The ALS is an efficient system which can deliver very dense and accurate point clouds from the ground surface and the objects which are located on it. Providing high quality height information of the landscape by means of LIDAR systems opens up an extensive range of applications in different subjects in photogrammetry and remote sensing. Moreover, laser scanning data is useful for an increasing number of mapping and GIS data acquisition purposes, including the detection and modeling of $3D$ objects. There are different types of information returned from the target which provide valuable information of the object and structures around it. Laser pulses have one important advantage: They partially penetrate the vegetation in gaps between leaves and thus make available data reflected from points underneath the vegetation. This property of the laser ray is at the heart of the difference between first- and last-pulse data: while in first-pulse data the vegetation's surface is represented well, this is not the case in last-pulse data.
Besides this, there are other important reasons why ALS systems are used for terrain modeling, with some of these reasons explained in the following:

- Day and night operation: ALS systems use active sensor systems illuminating the surfaces with their own energy, i.e., do not depend on reflected sunlight, as opposed to passive sensors such as cameras.

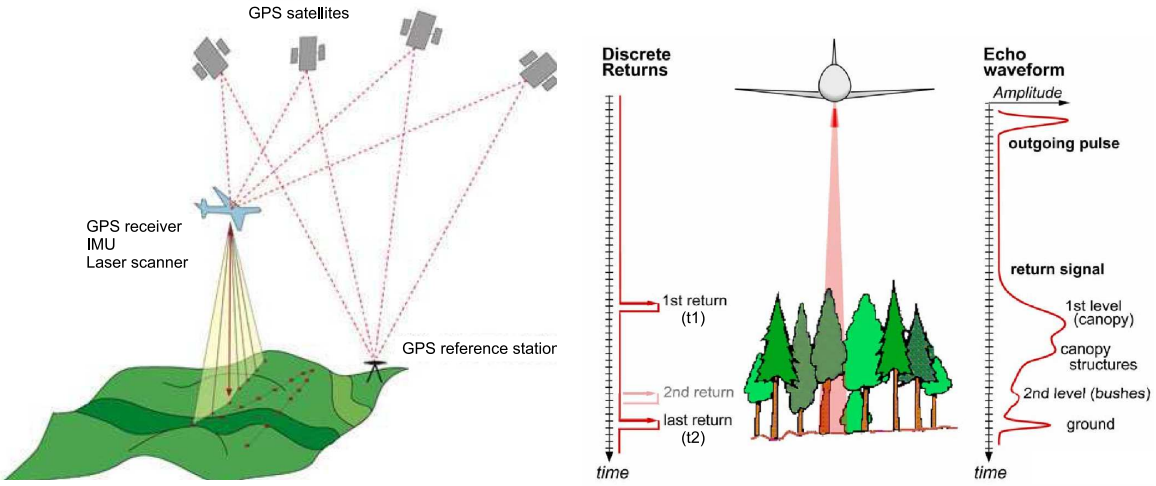
- Reduced weather dependency: ALS data collection is much less affected by weather conditions compared to aerial photography. Although the laser beam does not penetrate clouds well, flying at low altitude below the clouds makes it possible to still acquire data.

- Dense data collection: A very high density of height points is another significant characteristic of ALS. Since 1993 when Optech, a Canadian laser scanning system provider (Optech, 2007), delivered the first commercial laser scanning system to TopScan GmbH (TopScan, 2007), this young technology has developed significantly. The first system was able to transmit 2000 pulses per second (2 kHz), while the current system from this company sends up to 100.000 pulses per second (100 kHz). This increase directly affects the density of the ground points: the density of range points has improved to up to about 20 points per m^2 for a flying height of about 900 meter.

- High point accuracy: The height precision of an individual ground point is often in the order of $10 - 15cm$ (Maas, 2003).
- Fast data production and processing: ALS allows for an extremely rapid topographic data collection. With current commercial systems it is possible to collect one thousand square kilometers in less than 12 hours, having the geo-referenced DTM data available in 24 hours. The data collection process is completely digitally and therefore only few fieldworks are required.
- Operation on aircraft or helicopter

In summary, the rapid collection of highly accurate data, automatic data processing, and low operation costs compared to traditional surveying are advantages of ALS. To provide a precise terrain model, additionally to measuring the point heights it is necessary to determine accurately both the position of the aircraft and the orientation of the laser beam at the moment the distance is measured. Therefore, the following components have to work simultaneously for the generation of a precise digital surface model (cf. Figure 2.1(a)):

1. *Laser Range Finder (LRF)*: Measures the distance very accurately. It comprises the laser, transmission and receiving optics, the signal detector, the amplifier and the time counter.
2. *Scanner*: Deflects the laser beam across the flight path.
3. *Global Positioning System (GPS)*: Determines the position and path of the aircraft using differential GPS positioning.
4. *Inertial Measurement Unit (IMU)*: Measures acceleration and attitude changes and integrates them.



(a) Major components of an ALS

(b) Multiple return laser measurement (courtesy of IGI mbH, http://www.igi-systems.com)

Figure 2.1: Airborne laser scanning

A laser scanner is mounted in an aircraft in a similar way as an aerial camera. It sends an infrared laser signal towards the ground, which is then reflected back to the instrument. The

travel time to the ground and back to the instrument is recorded accurately. Pulses reflected by objects closer to the aircraft return earlier than pulses reflected by objects further away. As illustrated in Figure 2.2, the control unit of the laser scanning system triggers a pulse, the pulse is reflected from the surface and a part of the energy is scattered back to the receiver. Different pulse detection techniques can be used to estimate the time of flight of a pulse (Stilla and Jutzi, 2008).

Because of the path of the signal to the ground and back to the sensor, the recorded time (t) is divided by 2 and multiplied by the speed of light (C) to obtain the real distance (R) from the aircraft to the object.

$$R = \frac{C.t}{2}, \tag{2.1}$$

Most laser scanners utilize an oscillating mirror fixed in front of the laser, which deflects the

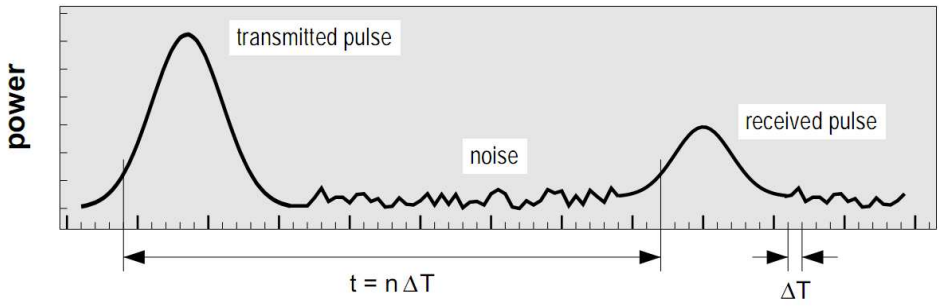


Figure 2.2: Principle of the distance measurement between the laser system and the surface. The time (t) is measured by counting the time intervals (ΔT) between characteristic points of the transmitted and received pulse. Range (distance) is computed by $R = \frac{C.t}{2}$ (Schenk, 2007).

beam in a zigzag movement orthogonally to the direction of flight.
The laser signal that is sent towards the ground may hit more than one object (Figure 2.1(b)). E.g., the first part of the signal may strike the top of vegetation canopy (t_1) while the last strikes the ground (t_2). In this case the canopy height (h) is determined by (cf. Figure 2.1(b)):

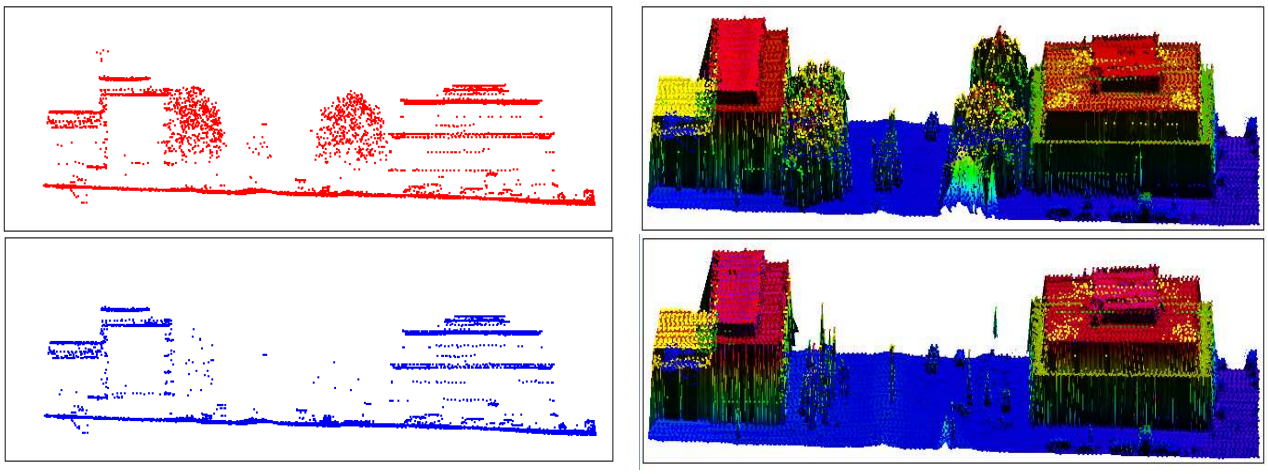
$$h = \frac{C}{2}(t_2 - t_1), \tag{2.2}$$

Depending on the system, the sensor records both pulses simultaneously. These two pulses are generally called First pulse (the portion which hits the foliage) and Last pulse (the portion which hits the ground).

Besides the range, some systems also measure the amplitude of the reflected signal or the entire signal over time (Jutzi and Stilla, 2006). From these information an intensity image is generated. The intensity of the received signal depends on the surface reflectance properties. If the distances to the first- and the last-pulse are recorded, then the intensities of the first- and last-pulses can be recorded as well.

2.2 Range Images and Range Features

One of the important and challenging problems in computer vision is the recognition of 3D objects. ALS systems provide depth values directly and, therefore, the raster image provided



(a) Side view (b) 3D view

Figure 2.3: First-pulse (top) vs. last-pulse (bottom) range data

from laser point clouds is called "range image". The gray values in a range image correspond to depth or height directly rather than to reflected light. Brighter gray values in a range image show that it is closer to the sensor or viewer. An important issue in any clustering and classification approach is the decision, what features should be used to distinguish defined classes. One of the helpful features which can be derived from range image is the surface normal or orientation in each point of the surface. The 3D laser points and surface normal information can be used for segmentation of the range image into "surface patches".
The other candidate features are the original range data (spatial coordinates x, y and depth value z), image gradients or local change, and curvature measures. Each feature descriptor plays a role for a better distinction between objects in image segmentation. The x and y features are important to provide connected segments. The height or depth value z help to detect "jump" edges, while the normal vector is needed to detect "crease" edges. In range images, jump edges occur where z values are discontinuous and crease edges correspond to points where surface normals are discontinuous (Jain and Dubes, 1988).

It has to be mentioned that further features of a reflected pulse like pulse width, shape of the pulse, number of reflective pulses, etc. can be exploited in case of full waveform recording (Stilla and Jutzi, 2008).

2.2.1 Gradient Image

For a continuous function $z = F(x, y)$ with two variables x and y the gradient vector is defined as:

$$grad(F) = (G_x, G_y) = (\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y})^T$$

$$|grad(F)| = \sqrt{(\frac{\partial F}{\partial x})^2 + (\frac{\partial F}{\partial y})^2}$$

With:
G_x: gradient in horizontal direction
G_y: gradient in vertical direction

The gradient is a vector and its magnitude characterizes the rate of height change per unit distance in the direction of the vector (SCHENK, 2004). In image processing the gradient image is usually calculated by convolving the image with a horizontal (or vertical) kernel:

$$G_x = F(x,y) * \frac{1}{2}\begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \tag{2.3}$$

$$G_y = F(x,y) * \frac{1}{2}\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{2.4}$$

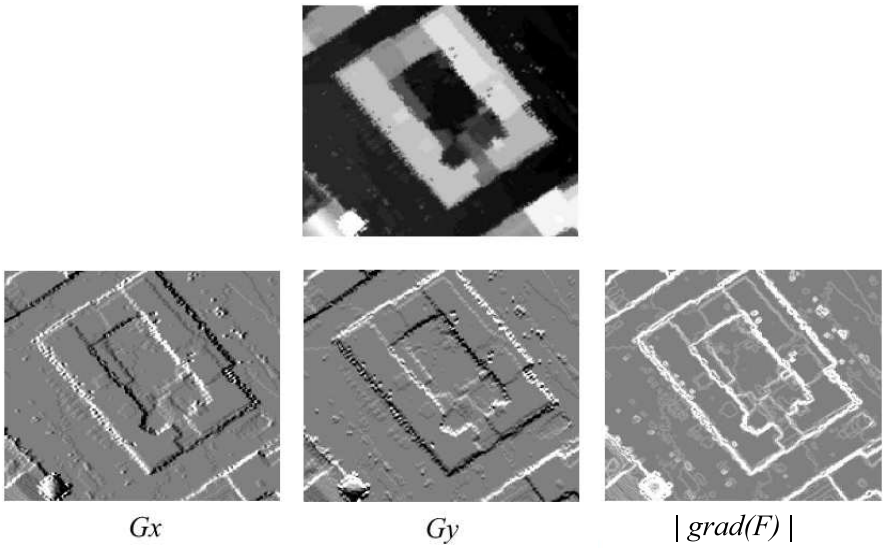


Figure 2.4: Gradient images

The problem with the above gradient computation is, that it is very noise sensitive. Therefore, in practice Sobel filtering, i.e., convolving image with a binomial filter before computing the difference is employed. The operator consists of a pair of 3×3 convolution kernels as shown in Figure 2.5.

+1	0	-1
+2	0	-2
+1	0	-1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Figure 2.5: Sobel filter kernels

2.2.2 Surface Normals

The surface normal is a vector perpendicular to a surface. The normalized normal vector or unit normal vector represents the orientation of a surface at a pixel. It can be estimated by

determining the best fitting plane over a small neighborhood (Jain and Dubes, 1988). A normal vector can also be computed by means of the cross product of any two non-collinear vectors that are tangent to the surface at the desired pixel. In each pixel in the image $F(u, v)$ there is exactly one tangential plane with its unit normal vector provided by:

$$\mathbf{n} = [\mathbf{n}_x, \mathbf{n}_y, \mathbf{n}_z] = \frac{X_u \times X_v}{\|X_u \times X_v\|} \tag{2.5}$$

If the embedding of the surface in $\mathbf{R}^3$ is given by $X = \begin{bmatrix} u & v & F(u,v) \end{bmatrix}^T$, then $X_u = \begin{bmatrix} 1 & 0 & F_u \end{bmatrix}^T$, $X_v = \begin{bmatrix} 0 & 1 & F_v \end{bmatrix}^T$, and accordingly:

$$\mathbf{n} = \frac{\begin{bmatrix} -F_u & -F_v & 1 \end{bmatrix}^T}{\sqrt{1 + F_u^2 + F_v^2}} \tag{2.6}$$

2.2.3 Gaussian and Mean Curvature

Curvature is a useful feature for classifying surfaces. The most important features which characterize the curvature of a surface are the mean and the Gaussian curvatures. Curvature basically expresses the rate of change of the unit normal vector to a surface with respect to the vectors in the tangent plane to the surface (Besl, 1988). The Gaussian curvature (K) and the mean curvature (H) can be expressed as follows:

$$H = \frac{X_{uu} + X_{vv} + X_{uu}X_v^2 + X_{vv}X_u^2 - 2X_uX_vX_{uv}}{2(1 + X_u^2 + X_v^2)^{3/2}} \tag{2.7}$$

$$K = \frac{X_{uu}X_{vv} - X_{uv}^2}{(1 + X_u^2 + X_v^2)^2} \tag{2.8}$$

In these equations, X_u and X_v are the first order partial derivatives and X_{uu}, X_{vv} and X_{uv} are the second order partial derivatives of the image. Based on the signs of the mean (H) and the Gaussian curvature (K) the eight basic shapes of the surface can be distinguished. These eight shapes are displayed in Figure 2.6.

2.3 Morphological Reconstruction Using Geodesic Dilation

2.3.1 Basic Operations of Gray-scale Mathematical Morphology

Mathematical morphology is employed for the analysis of spatial structures. Morphological operators are used for the selective extraction or suppression of image structures (Jähne et al., 1999a). The selection is based on shape, size, and orientation. It is achieved by probing the image with the structuring element (SE), which is a data set describing the given shape. A morphological operation transforms an image by means of a given structuring element into a new image, where structures of interest are emphasized.
Figure 2.7 depicts common operations of morphology with their mathematical definitions as well as examples which show the outcome of each operation on the same area. To compare

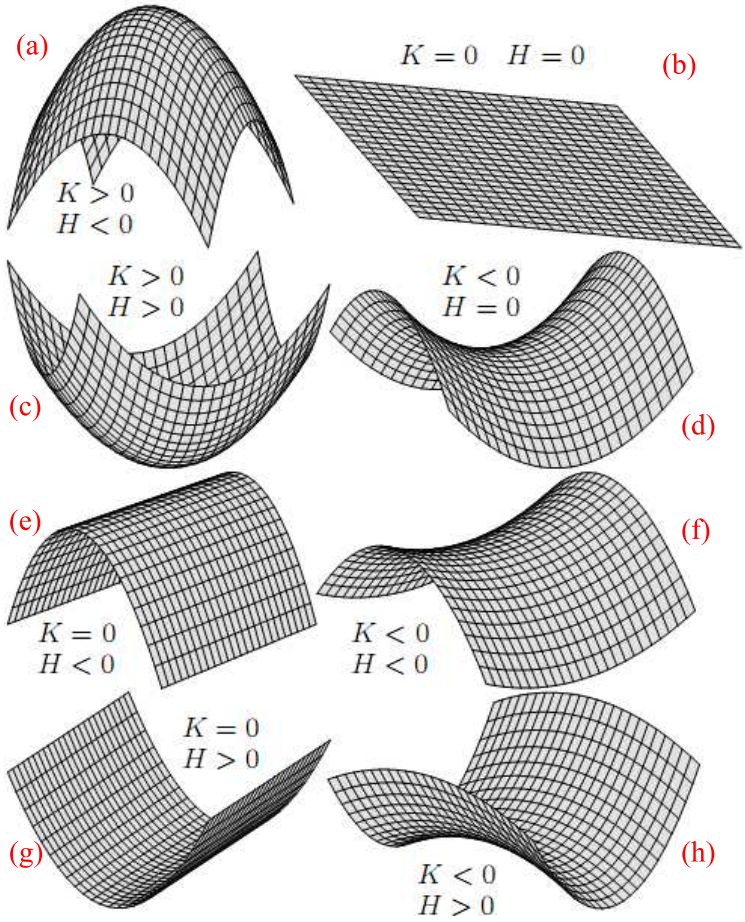


Figure 2.6: Characterization of the local shape of a surface by Gaussian mean and curvature: (a) peak surface, (b) flat surface, (c) pit surface, (d) minimal surface, (e) ridge surface, (g) valley surface, (h) saddle surface

the results, a profile is generated for each operation along the column marked by green in the "operation" column.

Erosion ($\varepsilon_B(X)$) and Dilation ($\delta_B(X)$) of X with structuring element B are the basic morphological operations and almost all other operations are generated based on these two. Erosion is defined as the location of points x, such that B is included in X when its origin is placed at x.

$$\varepsilon_B(X) = \{x | B_x \subseteq X\} \tag{2.9}$$

Equation 2.9 represents the erosion operation in binary image and X denotes the set of pixels which have values of 1. The definition can be extended to gray scale images: erosion of an image f by B is denoted by $\varepsilon_B(\mathbf{f})$ and is defined as the minimum of the translations of f by the vector $\boldsymbol{-b}$ of B or in other words, the eroded value at a given pixel x is the minimum value of the image in the window defined by the SE when its origin is at x (JÄHNE et al., 1999a):

$$[\varepsilon_B(f)](x) = \min_{\boldsymbol{b} \in B} f(x + \boldsymbol{b}) \tag{2.10}$$

Erosion is usually employed to eliminate unnecessary details and small objects from images. Accordingly, Dilation is the location of points x such that B hits X when its origin coincides

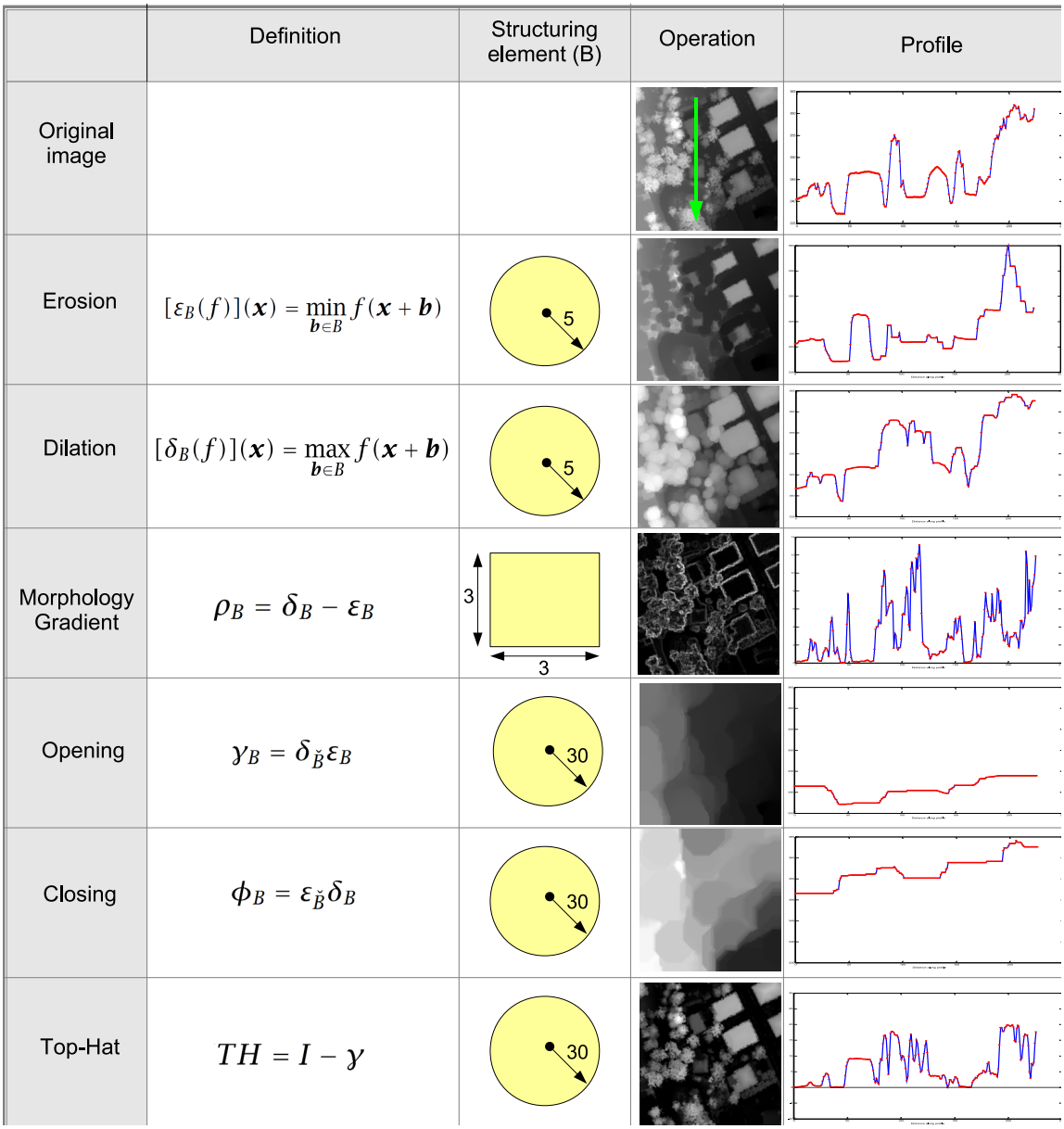


Figure 2.7: Common operations of mathematical morphology. The profile to the right stems from a column marked by a green marker in the "Operation" column.

with x (JÄHNE et al., 1999a). Therefore, binary and gray scale image dilation can be formulated respectively as:

$$\delta_B(X) = \{x | B_x \cap X \neq \varnothing\} \tag{2.11}$$

and:

$$[\delta_B(f)](x) = \max_{\boldsymbol{b}\in B} f(x+\boldsymbol{b}) \tag{2.12}$$

Contrary to Erosion, Dilation is an operation that grows or thickens objects in an image.
The morphological Gradient (ρ_B) is defined based on Erosion and Dilation. It is produced by subtracting the Erosion from the Dilation, with both operations using only one elementary SE. A square shaped SE of size 3 is employed in the example. The morphological gradient describes the local maximum variation of the gray level intensities in the image. It can be used

to describe local height jumps in range images (cf. Figure 2.7) highlighting the edges of $3D$ objects such as buildings and trees in LIDAR images.
Erosion shrinks objects and removes small ones. With Dilation with transposed SE the size of the remaining objects is reset to the original size. This forms Opening γ by a structuring element B and is denoted by γ_B (cf. Figure 2.7). Opening removes all pixels that are not covered by the SE when it fits the object pixels. In contrast to Opening, Closing (ϕ_B) fills all background structures that do not contain the SE.
As shown in Figure 2.7, a disk-shaped structuring element with a radius of 30 pixels is used to filter 3D structures in the image by Opening. The selection of proper SE is essential for the success of morphological operations. A SE should be selected based on knowledge about the shape, size, and orientation of the structures which are to be filtered. For the example in Figure 2.7, a SE with smaller radius eliminates small structures such as trees and small buildings. On the other hand, selecting a big SE may eliminate part of the hills in hilly areas as well as other large 3D structures. Therefore, an appropriate SE must be selected carefully.
The Top-Hat filtered image is calculated by an arithmetic difference between the image and its opening (cf. Figure 2.7). The Top-Hat extracts the objects that have not been eliminated by the opening. Objects larger than the structuring element are removed. Top-Hat is used to extract components with a large contrast with respect to the background. Typically, it removes slow trends and thus performs a contrast enhancement. Through the Opening structures smaller than the SE are removed thus the Top-Hat filtered image highlights those structures and prepares them for further processing (cf. Figure 2.7). For image with objects brighter than the background, the Top-Hat transform can be used for reducing the illumination gradient. A Top-hat with a large SE acts as a high-pass filter. The background trend lies within the low frequencies of the image and it can be removed by the Top-Hat (Jähne et al., 1999a). This can be understood by looking at the provided profile.
The main applications of the Top-Hat are:

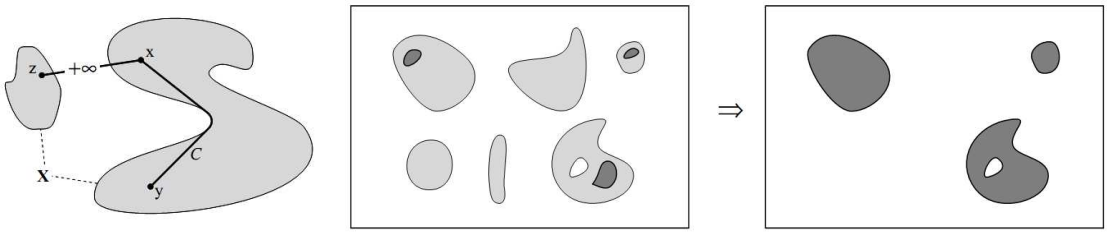
- The image contains inhomogeneous illumination.
- The image objects all have the same contrast and they are either darker or brighter than the background. If the object is darker than background, Closing should be employed instead of Opening.

2.3.2 Gray-scale Reconstruction Based on Geodesic Dilation

This section focuses on another morphological technique based on the geodesic distance. It is powerful, easy to implement, and suitable for a wide range of applications. All morphological operators discussed above involve a combination of one input image and an SE. Morphological operations based on geodesic distance employ two input images, namely *mask* and *marker* image. They allow the isolation of certain features within an image based on the manipulation of the *mask* and the *marker* image (Whelan and Molloy, 2001). Morphological reconstruction is based on the iterative application of erosions or dilations of the *marker* image until stability is reached, where propagation or shrinking of the *marker* image is perpetually limited by the image *mask*.

As illustrated in Figure 2.8(a), the geodesic distance between two points x and y within a given set X (the *mask*) is the length of the shortest paths joining x and y included in X (Vincent, 1993, 1997). According to this definition geodesic dilation and erosion are specified.

A geodesic morphological operation proceeds as follows: a basic morphological operator such as dilation or erosion is applied to the first input image, the *marker*. Then it is forced to remain



(a) Geodesic Distance between two points

(b) Binary reconstruction from marker

Figure 2.8: Image reconstruction of binary imagery based on geodesic distance (VINCENT, 1997)

either higher or lower than the second image, the *mask*. There exist two requirements: First, *mask* and *marker* must have the same size, and second, the *mask* image must have gray values higher than or equal to the *marker* image for geodesic reconstruction based on morphological dilation. In geodesic dilation the marker image is dilated by an elementary isotropic SE and the resulting image is forced to remain below the mask image. This means that the mask image acts as a limit for the dilated marker image. This is called geodesic dilation of size one. The geodesic dilation of size one of the marker image J with respect to mask image I is thus defined as (VINCENT, 1993):

$$\delta_I^{(1)}(J) = (J \bigoplus B) \bigwedge I, \tag{2.13}$$

In this equation $\bigwedge$ stands for the point-wise minimum between the dilated marker image and the mask image and $J \bigoplus B$ is the dilation of J with the elementary isotropic structuring element B. This equation indicates that the geodesic dilation of size 1 is generated by first computing the dilation of J by B and then selecting the minimum between the result and I at every point (x, y). The geodesic dilation of size n of the marker image J with respect to the mask image I is obtained by performing n successive geodesic dilations of size one of J with respect to I (GONZALEZ and WOODS, 2008):

$$\underbrace{\delta_I^{(n)}(J) = \delta_I^{(1)}(J) \circ \delta_I^{(1)}(J) \circ .... \circ \delta_I^{(1)}(J)}_{n-times} \tag{2.14}$$

Equation (2.14) also defines the morphological reconstruction by geodesic dilation of the mask I with respect to the marker J. The desired reconstruction is achieved by carrying out geodesic dilations until stability is reached. I.e., morphological reconstruction can be thought as repeated dilations of the marker image while the contour of the marker image fits under the mask image. By this means, the peaks in the marker image spread out, or dilate. Each successive dilation is forced to lie underneath the mask. Figure 2.8(b) shows the application of geodesic image reconstruction to extract the desired binary regions. Some seed regions (dark gray in Figure 2.8(b)–left) are initially selected and corresponding regions are then extracted by binary geodesic reconstruction.

Figure 2.9 presents the morphological reconstruction of a 1D signal g from a marker signal f. The 5-fold geodesic dilation of the marker signal with respect to the mask signal is equivalent to the reconstruction of g from f since further geodesic dilations do not change the result anymore. In this figure, (a) presents the original marker signal f and the mask signal g; (b) to (f) show the results of repeated geodesic dilations of f with respect to g.
Morphological reconstruction based on geodesic dilation has some unique properties compared to basic morphological operations (MATHWORKS, 2007):

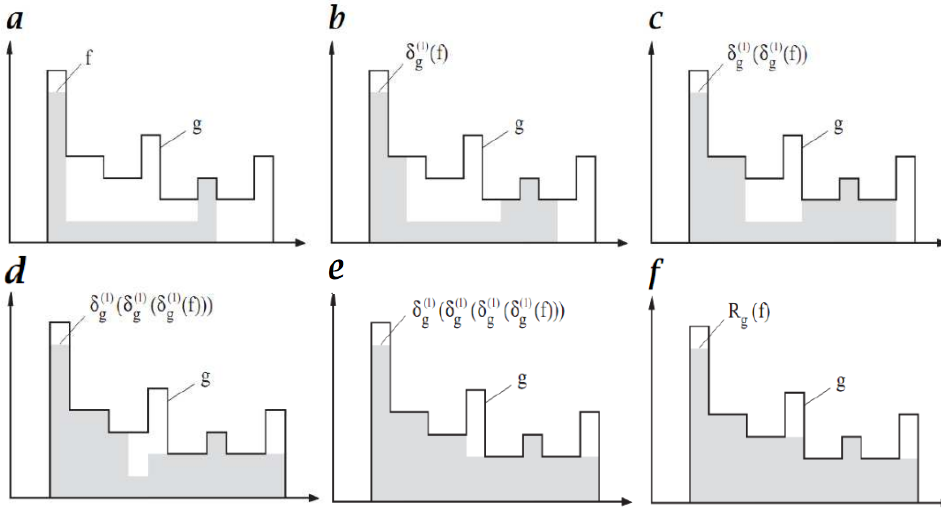


Figure 2.9: Morphological reconstruction by geodesic dilation; Geodesic dilation of size 5 of the marker f with respect to the mask g is equivalent to the reconstruction of g from f, because further geodesic dilations do no longer change the result (JÄHNE et al., 1999a)

- Two images are involved in processing, rather than one image and an SE.
- The processing is repeated until stability, i.e., until the image no longer changes.
- The procedure is based on connectivity rather than on an SE.

The method explained above and formulated in Equations (2.13) and (2.14) is the standard way to perform image reconstruction by geodesic dilation. I.e., combined basic dilation and point-wise minimum operations are iterated until stability to produce the reconstructed image.

Depending on the size of the image, the number of iterations increases and therefore, the procedure becomes very time consuming. VINCENT (1993) proposed a fast and efficient algorithm for gray scale image reconstruction by dilation which is termed as "Fast Hybrid Reconstruction Algorithm". The algorithm uses three First-In-First-Out (FIFO) operations:

- *fifo_ add(p)*: It places the pointer to pixel p into the queue.
- *fifo_ first()*: It returns the pointer to the pixel at the beginning of the queue and removes the pointer from the queue.
- *fifo_ empty()*: It returns *true* if the queue is empty and *false* otherwise.

Accordingly the "Fast Hybrid Reconstruction Algorithm" operates as follows:

- I : mask image
- J : marker image
- Scan J in raster order:
  - Let p be the current pixel;

  - $J(p) = (max\{J(q), q \in N_G^+(p) \cup \{p\}\}) \wedge I(p)$

- Scan J in anti-raster order:
  - Let p be the current pixel;
  - $J(p) = (max\{J(q), q \in N_G^-(p) \cup \{p\}\}) \wedge I(p)$
  - If there exists $q \in N_G^-(p)$ such that $J(q) < J(p)$ and $J(q) < I(q)$ then *fifo_ add(p)*

- Propagation: While (!fifo_ empty)
  - p = fifo_ first()
  - For every pixel $q \in N_G(p)$:
    * If $J(q) < J(p)$ and $I(q) \neq J(q)$ then
      $J(q) = \min\{J(p), I(q)\}$
      fifo_ add(q)

$N_G(p)$ denotes the neighborhood of p on a grid G, and N^+ as well as N^- the neighboring pixels of pixel p for raster and anti-raster scan filtering, respectively (cf. Figure 2.10). $\wedge$ is the point wise minimum operator (Vincent, 1993) .

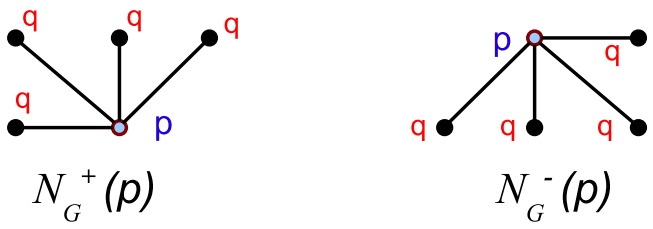


Figure 2.10: Neighboring pixels for the filtering of point p in raster (left) and anti-raster (right) order scanning in the hybrid geodesic dilation algorithm.

This "hybrid" method begins with *raster order* scanning, i.e., from left to right and from top to bottom. For each candidate pixel $J(p)$ the maximum value of it and the neighboring pixels $J(q)$ (cf. Figure 2.10 – left) is determined. The point-wise minimum between this maximum value and the mask image in point p, i.e., $I(p)$, is used to replace $J(p)$.

The process is repeated in *anti-raster order*, i.e., from right to left and from bottom to top. Every pixel p for which its current value could still be propagated, i.e., for which holds:

$$\exists q \in N_G^-(p),\ J(q) < J(p) \text{ and } J(q) < I(q),$$

is added to the queue. The elements of the queue are finally propagated by checking their neighborhood. For every points q in the neighboring of p it checks if q is lower than p and if it is necessary to propagate. Therefore, the minimum of the marker in p and the mask in q is replaced in $J(q)$. The first element is taken away from the queue until it is empty.

Some major applications of the morphological reconstruction for extracting and segmenting the prominent parts of the image are explained in the following:

1. Removal of holes in binary or gray scale images: In binary images holes are the connected components of the background that are not connected to the image border (JÄHNE et al., 1999a). Similarly, in gray scale image all minima pixels which are not connected to the image boundary can be regarded as holes. In ALS data these minima pixels are often outliers or noise and should be eliminated. Algorithms which are based on basic morphological operations such as Closing are not efficient and they may even produce new minima (JÄHNE et al., 1999b). A geodesic morphological operation is used to eliminate holes from the image. It can be constructed from geodesic erosion or geodesic dilation. For geodesic dilation for the *mask* image I the complement of the original gray scale or binary images is used and the *marker* image is generated as follows:

$$J(i,j) = \begin{cases} I(i,j) & \text{if point } p(i,j) \text{ lies on the border of } I \\ min(I) & \text{otherwise} \end{cases} \tag{2.15}$$

   In geodesic erosion the mask image is equal to the input image and and the marker image is an image of constant value $max(I)$ having the same border values as the input image (cf. Figure 2.11).

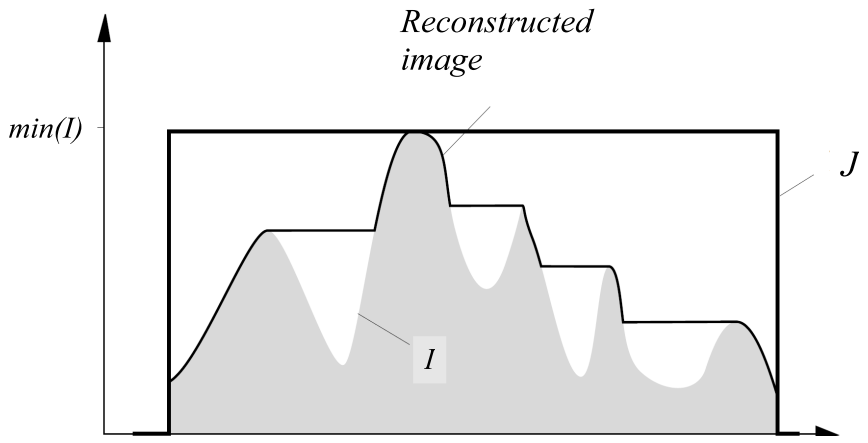


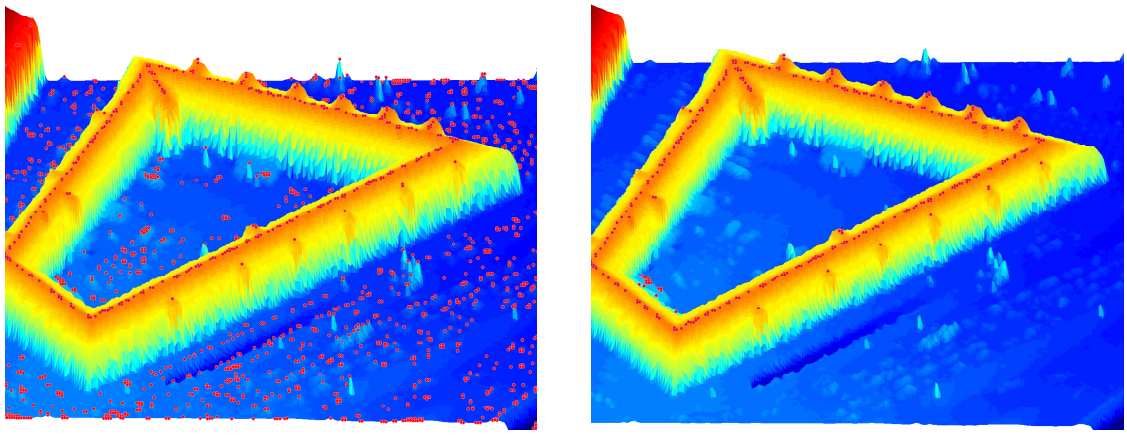
Figure 2.11: Filling holes on a 1-D signal I using marker J: all inner minima of I are removed by the morphological reconstruction of I from the marker function J.

2. Regional Extrema Extraction: Extracting regional extrema, i.e., minima or maxima, of an image is one of the important tasks in image processing. By definition (JÄHNE et al., 1999a), a regional minima M of an image I at elevation t is a connected component of pixels with the value t whose external boundary pixels have a value strictly greater than t. The regional maxima is the complement of it. It should not be confused with *local maxima* which are the pixels p for which holds: Their value $I(p)$ is greater than that of any pixel in their neighborhood. Both regional and local extrema can be extracted by means of geodesic morphology.
   Local extrema are usually limited to a few pixels in the image and do often represent the errors or outliers. Therefore, they should be eliminated from the data set before further processing because most of the image analysis operations are sensitive to noise and errors and they thus reduce the quality of the result.
   We focus on extracting local or regional maxima by geodesic dilation. Extracting local

or regional minima can be approached by geodesic erosion. The original image is selected as *mask* image ($I(i,j)$) and the *marker* image is provided by:

$$J(i,j) = I(i,j) - h$$

where h denotes an offset value. For the offset h often a small value is chosen, such as 1 in an 8-bit image. Figure 2.12 shows a practical application for regional maxima extraction



(a) All regional maxima in the area calculated using an offset $h = 10$ *cm*

(b) The points outside the building are eliminated based on the segmentation result

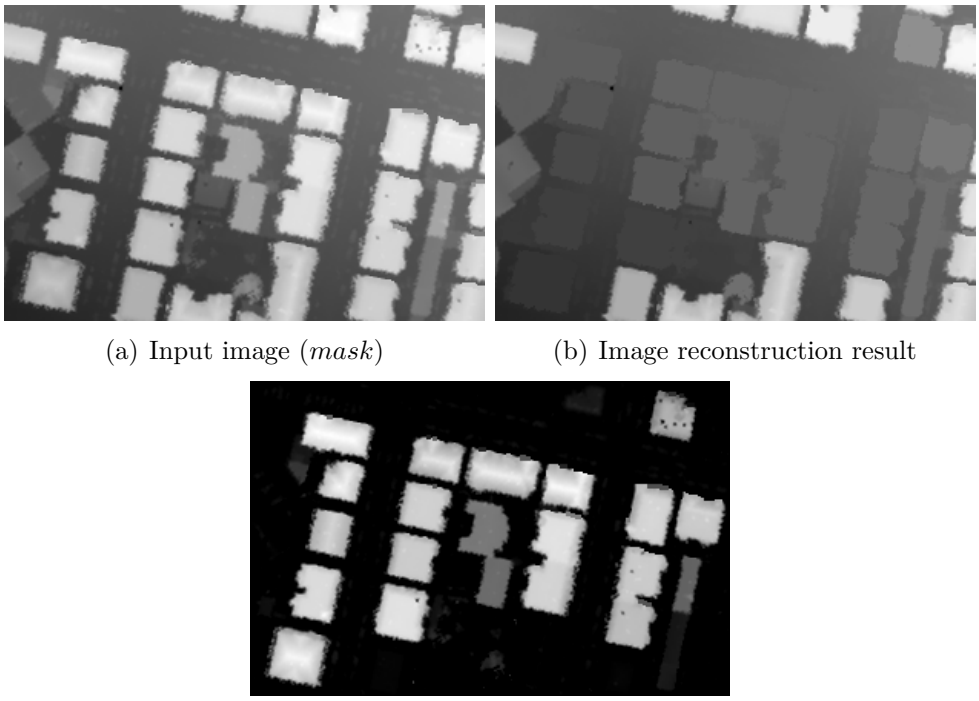
Figure 2.12: Application of geodesic morphology for extracting ridge points of a building

by means of geodesic morphology. A building with several gable roofs is located in the image. The ridge line is represented by the location of the regional maxima in the image. Figure 2.12(a) displays the location of all regional maxima calculated by employing the original laser image as *mask* and a *marker* image generated by subtracting 10 *cm* from the *mask*. Not only the location of the ridge points, but also all other regions such as the top of the trees and bushes are extracted. After the classification of the building the regional maxima points located outside the building can be eliminated. The internal points are used later for the classification of the ridge points.

3. Extraction of structures connected to the image border: In image processing it is often necessary to evaluate only structures that are completely located inside the image. Therefore, objects which are only partly inside the scene should be detected. This is done as follows: The input image is selected as *mask*. The *marker* image is produced in the form of an image with values equal to *zero* except for the border pixels which get the value of the *mask* image. Subsequently, to remove the pixels connected to the border, the reconstructed image must be subtracted from the input image. Figure 2.13 shows results for the elimination of objects connected to the border of a *LIDAR* image. The objects connected to the boundary are highlighted by performing image reconstruction as displayed in Figure 2.13(b). Finally, to eliminate those objects, the reconstructed image (cf. Figure 2.13(b)) is subtracted from the *mask* image (cf. Figure 2.13(a)) shown in Figure 2.13(c).

2.4 Robust Model Fitting

This section discusses the robust models employed in this thesis to fit straight lines and planes to point clouds. The *Hough Transform* (Hough, 1962) is one of the most common approaches



(a) Input image (*mask*)

(b) Image reconstruction result

(c) Elimination of the objects connected to the border

Figure 2.13: Extracting the objects connected to the border of image in $LIDAR$ image

to fit linear features such as straight lines in 2D space. It is applied here to fit straight lines to binary images. In 3D space *RANSAC – RANdom SAmple Consensus* (Fischler and Bolles, 1981) is utilized for fitting geometric models, i.e., straight lines and planes, to the data set.

2.4.1 Hough Transform

The Hough Transform is a technique which can be used to robustly extract lines, circles, and ellipses from images. It defines a mapping from the image points to a parameter space which is also called "Hough space". In this thesis the focus is on extracting 2D lines from binary images.
It is the assumed that the image is already segmented into binary regions including the linear regions. Consider a point with coordinate (x_i, y_i) located on a straight line formulated in slope-intercept equation form as $y_i = ax_i + b$ in the xy plane. To find the all points on this line, the line equation is reformulated as $b = -x_i a + y_i$ which is the equation of a single line in the so called "parameter space" in the ab plane. As illustrated in Figure 2.14, two points (x_i, y_i) and (x_j, y_j) which are located on the line in point space (left figure) represent a point (a', b') in the parameter space (middle figure), where a' and b' are the slope and the intercept of the line shown in the left figure. As the slope parameter a goes toward infinity for vertical lines, usually the normal representation of the line is used:

$$x \cos\theta + y \sin\theta = \rho$$

In this case the parameter space is formed by θ and ρ instead of a and b. Sinusoidal curves represent the location of all lines passing through a particular point (x_i, y_i) in the $\rho\theta$ plane (cf. Figure 2.14 right) (Gonzalez et al., 2004). In an accumulator array the values for θ and ρ

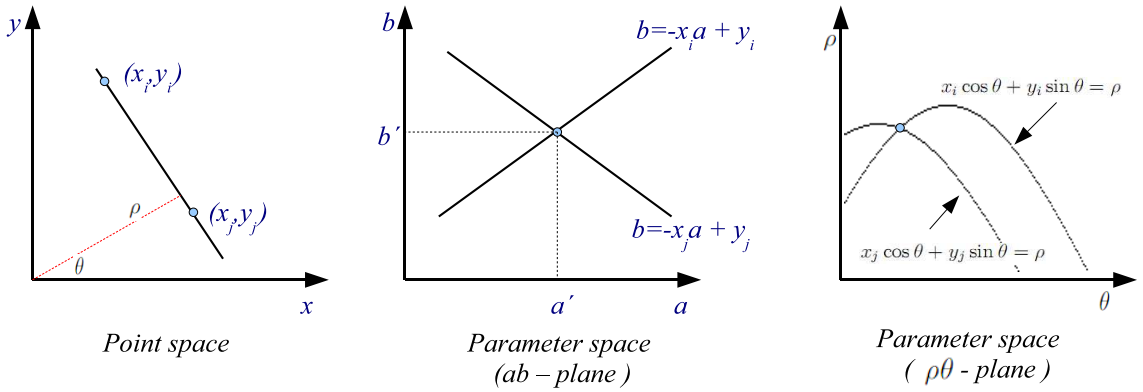


Figure 2.14: Representation of point and parameter spaces of the Hough Transform

are accumulated. Accordingly, the cell with the maximum value represents the strongest line in the image.

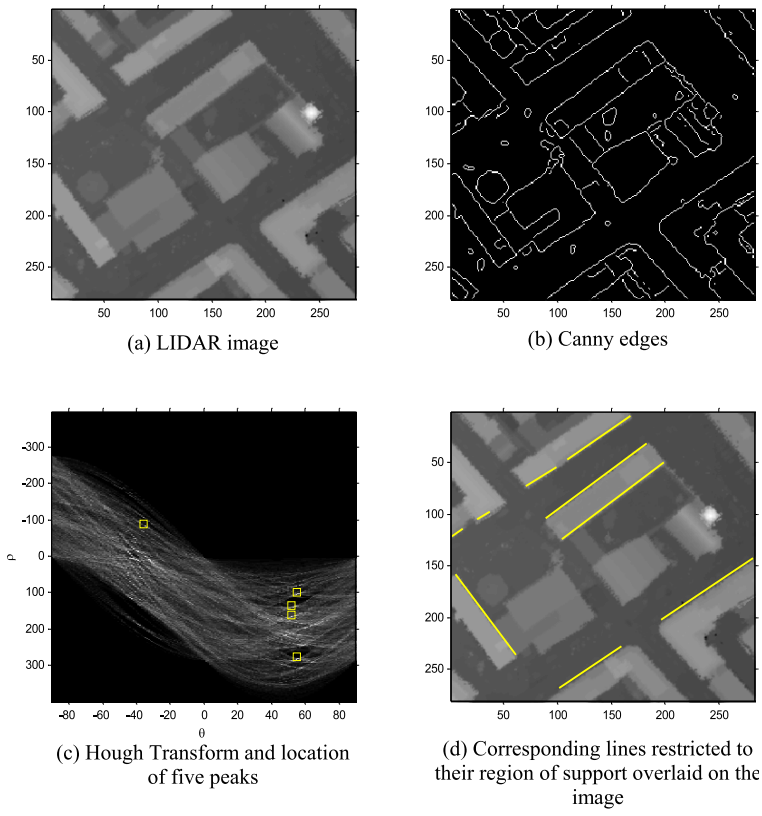


Figure 2.15: Line Extraction by Hough Transform

Figure 2.15 illustrates how the Hough Transform employed for straight line extraction from LIDAR image. Here, Canny edge detector is utilized to extract the binary edges of the image.

2.4.2 Random Sample Consensus

RANdom SAmple Consensus (RANSAC) is a method for the robust fitting of a model to a data set which contains many outliers (Fischler and Bolles, 1981). The basic assumption of RANSAC is that there are *inliers* and *outliers* in the data set. The inliers are the data

which do fit to the expected model and the outliers are the data which do not fit and come from incorrect measurements, noise or other sources. Figure 2.16 shows how RANSAC works

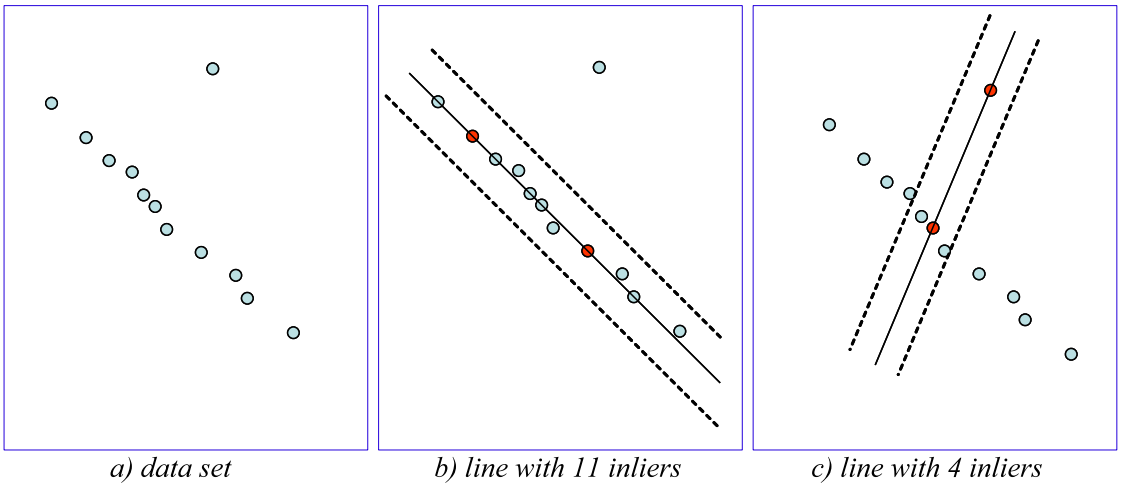


Figure 2.16: Robust line fitting with RANSAC

to fit a straight line to a data set. The input is a data set, a parameterized model, which is to be fitted to the data set, and the following confidence parameters which should be tuned beforehand (Forsyth and Ponce, 2003):

- Minimum number of points required to fit the model, i.e., sample size (n): This input data is fixed as soon as selecting parameterized model. E.g., 2 points is the minimum number of points required to determine the parameters of a straight line in 2D and 3 is the minimum number for plane fitting in 3D.
- Minimum number of iterations (k): The number of iterations, i.e., the number of samples, is usually selected large enough, to be sure with a probability of p, that at least one of the selected random samples is free of outliers. According to (Hartley and Zisserman, 2004), the minimum number of k can be calculated by:

$$k = \log(1-p)/\log(1-(1-\epsilon)^n) \tag{2.16}$$

  With:
  p: Probability, that at least one sample has no outlier
  ϵ: Probability, that any selected data point is an outlier
  n: Sample size

- The threshold used to identify a point that fits well, i.e., an inlier (t): This value defines the maximum acceptable distance between the point and the fitted model. If the distance is below the threshold, the point is accepted as an inlier. In general this value can be defined relatively simply from empirical knowledge or by experiment.
- The size of acceptable consensus set for termination of iteration (d): It is the criterion for the immediate termination of the iteration. I.e., if in one iteration step the number of d inliers is obtained, the iterative procedure will be terminated. This value can be derived for ϵ and m, the number of points in the data set as (Hartley and Zisserman, 2004):

$$d = (1-\epsilon)m$$

The algorithm for robustly fitting a model to a data set S containing outliers after tuning the above parameters is summarized follows:

1. Randomly select n data points from S.
2. Generate the mathematical model from the n points.
3. Determine the points S_i having a distance to the model smaller than or equal to the threshold t.
4. If the number of elements of S_i, i.e., number of inliers, is greater than d, re-generate the model using the whole data set S_i.
5. If the size of inliers is less than d, select a new random sample and repeat steps 2 and 4.
6. After k iterations of the steps 1 to 5, the largest consensus set of S_i is selected and the model is re-generated using all the points in S_i.

Figure 2.16 shows a RANSAC based method for fitting straight line to a sample data set containing 12 points. Two random points are selected (red points) and the line is fitted to them (continuous black line). The inliers of the lines are determined as points between two parallel lines (dashed black lines) defined by a distance threshold to the line. As shown the produced line in Figure 2.16(b) contains 11 inliers and 1 outlier while, the line in Figure 2.16(c) contains 4 inliers and 8 outliers and thus the former would be selected.

2.5 Building Models

This section outlines the terminology regarding the most commonly used building structures in the scene. The focus is on the geometric structure of roofs as well as the Levels-Of-Detail (LOD), which are defined for 3D representations of the building model.

2.5.1 Roof Types and Modeling of Buildings

In general, a building model is formed by combinations of basic building parts which are mostly characterized by their roof structure as shown in Figure 2.17. The smaller structures on top of the roof such as dormers and chimneys can be integrated with these basic models.

Figure 2.18 displays a building model which is formed by the combination of a hip roof and a gable roof. The figure shows the main structures of the building which can be used for building reconstruction. A brief description of the terms highlighted in the figure is given in the following:

- Ridge: The highest points of a roof, often in the form of a horizontal line
- Eave: The lower edge of an inclined roof
- Dormer: A vertical structure above the roof usually with a window
- Hip: The convex angle formed by intersecting two inclined roof parts
- Valley: The concave angle formed by intersecting two inclined roof parts

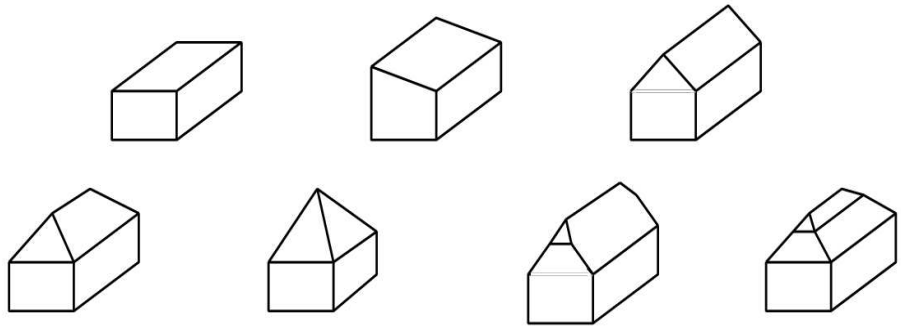


Figure 2.17: Basic building models - flat roof, pent roof, gable roof (first row) and hip roof, pavilion roof, hipped-gable roof, and mansard roof (second row) Müller (1998)

- Gable: The vertical triangular part of the roof from the cornice or the eaves to the ridge
- Gutter: The channeled component along the eaves used to collect and carry off rainwater
- Building outline: The outer boundary of a building which contains all structures attached to it such as house(s), garages and balconies. Attached trees or other vegetation are excluded. The building outline is also referred to as "footprint".

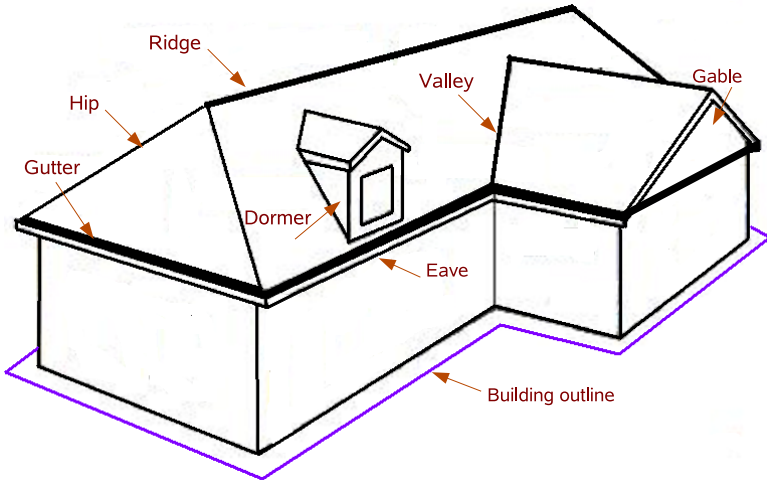


Figure 2.18: Structures of a roof

Buildings are available in real world in different shapes and models. Förstner (1999) distinguishes two different building models:

- Parametric models: A model is described by a set of parameters. A building with a gable roof defined by length, width, and two heights is such a parametric model. Other buildings with flat roof, gable, or hipped roof as well as L-shaped buildings can be represented as parametric models.
- Generic models: This allows for the variation of the basic building structure. Three main classes exist:

- Prismatic models: A Building consists of a polygonal ground plan, vertical walls, and a horizontal roof.
- Polyhedral models: This is more general but restricted to buildings for which the surface is formed by planar patches.
- Constructive Solid Geometry (CSG) models: Here, the building is formed by combining simple parts (primitives) using boolean operators such as union and intersection (Förstner, 1999).

All building models are generated by a combination of the above-mentioned basic models. In reality, these categorization is closely related to the methods used for modeling. The methods for building reconstruction are classified into two major classes: model-driven and data-driven. Generally, model-driven methods are related to parametric models which data-driven methods are related to generic models. A limited number of building models are predefined in model-driven approaches. Model-driven implies top-down approaches which begin with a model as hypothesis and then use data to verify the model. This can be easily implemented, but have a limited application in practice, because they can only model simple buildings, such as flat-roof and gable buildings (Ma, 2004). Data-driven means bottom-up modeling, which begins with the extraction of building primitives, e.g., building faces. After analyzing the surface topology, the building model is constructed. In practice, this approach can handle more buildings than the model-driven approach.

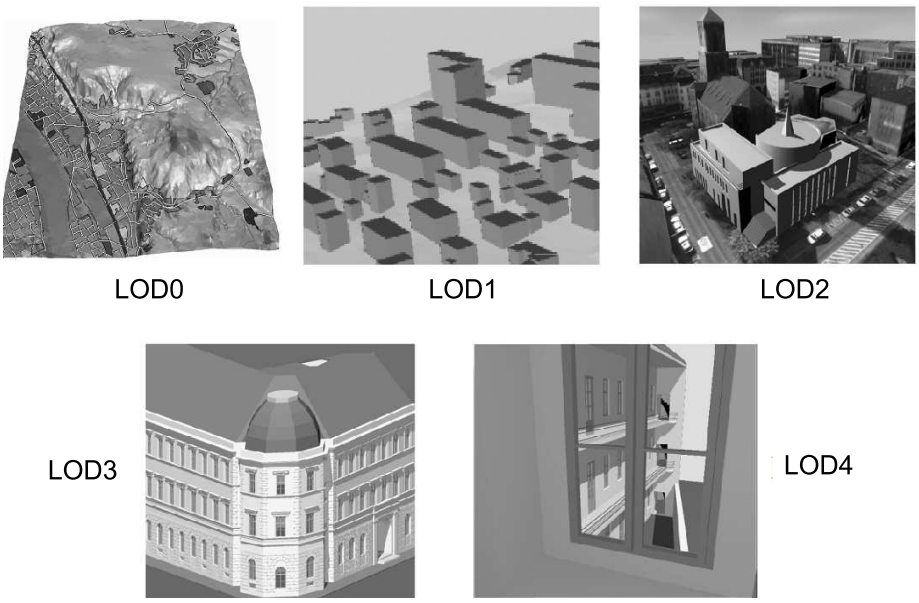


Figure 2.19: The Levels-of-Detail defined by CityGML (Kolbe et al., 2005)

2.5.2 Levels of Detail

A building model can be represented in 3D including a different amount of detail information. Different levels-of-detail (LOD) are defined in CityGML (Kolbe et al., 2005) which categorizes geometrical, topological, and semantic information into five LODs (cf. Figure 2.19). The categorization begins with LOD0. It is the coarsest level and it contains a 2.5D Digital Terrain Model (DTM). In this level an aerial image can be overlaid on the DTM. The LOD1 is a

building block model not including any roof structure or textures. Roof structures as well as the building textures are included in LOD2. LOD3 is generated by adding architectural information with detailed walls and smaller structures of the roofs such as balconies, bay windows, and projections. Finally in LOD4, interior structures such as rooms, stairs, and even furniture are added to the LOD3.

Chapter 3

Former Work

Automatic 3D building modeling, in general, comprises two major steps, namely building detection and 3D reconstruction. The procedure for LIDAR data processing consists of DTM generation or filtering of non-ground regions, building detection, and building reconstruction. Many studies have been carried out to use airborne laser data as well as other data such as aerial imagery or satellite images for the generation of 3D building models. In this chapter an overview over former work on the processing of image and ALS data for the generation of 3D building models is given.

3.1 3D Building Reconstruction from LIDAR Data

The overview of the former work on the generation of building models from ALS data is divided into three parts: filtering of the non-ground regions to generate the bare earth, building outline extraction as well as approximation, and finally 3D building modeling.

3.1.1 Filtering of Non-ground Regions and DTM Generation

ALS has become an accepted data source for the highly automated acquisition of digital surface models (DSM) as well as for the generation of digital terrain models (DTM). To generate a high quality DTM from LIDAR or DSM data, 3D non-ground points have to be separated from the ground points. Various techniques and filtering methods have been proposed to generate DTM from LIDAR or other DSM data even though one has to note that the research in this field is still ongoing.
In early work Kilian et al. (1996) presented first ideas on how to generate a DTM from LIDAR data recorded in wooded areas based on gray-scale morphological opening (cf. Section 2.3.1). The lowest points within a window of a given size are first detected by opening. Then the points in this window that fall within a band above the lowest elevation are considered as ground points and a weighted surface interpolation using these points is applied to compute the DTM. A conclusion of this work was that the size of the Structuring Element (SE) used for opening is a critical parameter for which there is no single optimal value. Therefore, the use of multiple openings with different sizes for the SE was suggested.

Kraus and Pfeifer (1998) introduced another algorithm for DTM generation in wooded areas based on linear prediction. They also start with an approximation of the ground surface. The vertical distances from the ground surface to the approximated surface are used to define

weights which are employed in computing the DTM based on linear prediction. Points that are above the surface obtain a small weight and those are below the surface get a large weight. The new surface is then recomputed using linear interpolation function and the corresponding weights. The procedure is continued until a certain number of iterations have been achieved or the computed surface does not change considerably comparing to the previous surface.

Axelsson (1999) described a method for DTM generation based on the progressive densification of a triangular irregular network (TIN). The idea is to fit a surface from below to the point cloud. In every iteration the surface is allowed to fluctuate within a certain range and points from the point cloud are added to the TIN. It is iterated until no further low ground points can be added. The approach has been implemented in the TerraScan software package (Terrasolid, 2007).

Vosselman (2000) proposed a slope based filtering method for separating non-ground points from ground points. A point is classified as a terrain point if there is no other point in its vicinity to which the height difference is larger than an allowed maximum difference. The method assumes that the terrain slopes do not exceed a certain threshold and the features having slopes above this threshold are non-ground objects. The assumption limits the approach to terrain with gentle slopes. A modified slope based filtering was proposed based on varying the threshold value according to the terrain slope (Sithole, 2001).

Wack and Wimmer (2002) proposed a hierarchical grid-based approach for the generation of DTMs from airborne laser data. They start with a coarse grid of $9m$ width and define the raster height by selecting the height within the raster element for which 99% of all points are higher. The Laplacian of Gaussian (LoG) operator in combination with a weight function is utilized to detect and remove points that are not considered to be ground points.

A progressive morphological filtering method was developed by Zhang et al. (2003) with the focus to remove the non-ground measurements from LIDAR datasets. The algorithm utilizes morphological opening and gradually increases the size of the SE. The resulting elevation differences are used to classify ground and non-ground points by applying a threshold which depends on the SE size. An unsupervised statistical skewness balancing segmentation algorithm is proposed by Bartels and Wei (2006) to separate non-ground regions from the ground regions. It is based on the assumption that the naturally measured samples lead to a normal distribution and non-ground objects disturb this distribution. A recursive point removal based on the skewness is utilized until a normal distribution is achieved.

Stilla and Jutzi (2008) have been shown that full-waveform analysis enables extraction of more information compared to classical analogous pulse detection methods. The shape of the pulse and the entire signal can be considered for determining the range more accurately. Further improvements on reliability and accuracy can be derived by signal-processing methods based on the transmitted and the received waveform, e.g., deconvolution. Additionally, attributes of the surface can be derived from a parametric description of the waveform. The attributes maximum amplitude and pulse width support the discrimination between volume scatterer (vegetation) and hard targets (man-made objects) (Kirchhof et al., 2007). For filtering points belonging to buildings, the introduction and test of hypotheses about the shape of the surface (e.g., plane, sphere) may efficiently support the analysis of single waveforms. Two different strategies assuming a planar shape in the local neighborhood of the surface and introducing this assumption into the signal analysis should be addressed. Both strategies combine information from top-down (surface primitives) and bottom-up (signal processing) for an extended analysis of full waveform laser data. The first strategy (Kirchhof et al., 2007) uses an iterative processing of waveforms considering a predicted shape of the waveform from the local neighborhood. A pre-segmentation based on surface attributes is carried out to distinguish between partly penetrable

objects (e.g., trees, bushes) and impenetrable surfaces (e.g., roof, wall). Derived range values from presegmentation of the impenetrable surfaces are used to automatically generate surface primitives (e.g., planes). This allows a refinement of each range value, considering the surface geometry in a close neighborhood. Furthermore, partly occluded surface areas are extended by prediction of the expected range values. This prediction is further improved by considering the surface slope for the estimated received waveform. Expected pulses are simulated and correlated with the received waveforms. Accepted points that were missed in the first processing step due to weak signal response are associated to the point cloud. The procedure is repeated several times until all appropriate range values are considered to estimate the surface. The second strategy Stilla et al. (2007) uses a slope compensated stacking of waveforms. Weak pulses with a low SNR are discarded by classic threshold methods and get lost. In signal and image processing, different stacking techniques are used to improve the SNR. For detection of weak laser pulses, hypotheses for planes of different slopes (e.g., angle difference 5^o) are generated. According to the slope of the hypothesis, the waveforms in the local neighborhood are shifted in range. A superimposed signal is calculated from the stack of shifted waveforms. The maxima of superimposed signals from all hypotheses are compared to verify a hypothesis. Each signal is assessed by a likelihood value with respect to its contribution to the accepted hypothesis. Finally, signals are classified according to the likelihood values obtained using two thresholds and visualized by the traffic-light paradigm. The results contain detected pulses reflected from objects, which cannot be predicted by the previously detected point cloud.

3.1.2 Building Outline Detection and Approximation

Building detection from LIDAR data has been an intensively studied subject over the past few years. Almost all algorithms for building reconstruction initially have a need for the location of the building in order to start the reconstruction. This initial information can be extracted by digitizing 2D maps or by automatic segmentation and classification. A number of different techniques have been proposed for segmentation of ALS range or DSM data with the aim of building extraction. They are based on the assumption that buildings are higher than the neighboring topographic ground surface.

A simple thresholding with the threshold value t computed by analyzing the histogram of height values can be used for segmentation in very special cases. In regions with steep terrain, high vegetation, and vegetation close to buildings this approach is not applicable (Maas, 1999b). Most existing building extraction methods use last pulse ranges to generate DSM and DTM in the first stages (cf. Section 3.1.1). The normalized DSM (nDSM) is then produced by subtracting these two data sets. The non-ground regions, mainly consisting of buildings and trees, are then extracted using height thresholding.

Weidner and Förstner (1995) proposed DTM generation based on DSM and morphological processing. A global threshold is applied to the nDSM for the segmentation of 3D regions. For discriminating buildings from vegetation objects, two features, one based on step edges and the second using the variance of the surface normals, have been employed.

The analysis of height texture for automatic segmentation and classification of dense ALS data (more than 5 points per m^2) is proposed by Maas (1999b). The original height data and two height texture features generated from the laser data are used as basis of a maximum Likelihood classification. The two texture features consist of the Laplacian filtered image and the maximum slope around each pixel. An overall accuracy of about 98% is reported by Maas (1999b). Recording first-pulse as well as last-pulse laser data was reported to produce very useful information for the discrimination of building and vegetation pixels.

Rieger et al. (1999) used three image bands namely normalized first-pulse, normalized last-pulse, and vegetation index. The vegetation index image is generated by subtraction of first- and last-pulse images. A simple height threshold was employed to discriminate between buildings and vegetation assuming that all pixels having heights less than $1m$ in the vegetation index image belong to buildings. This assumption is not always true because (especially high) building boundaries often lead to high values in the vegetation index image with heights of more than a few meters. The other way around, in some regions the trees in both the first- and last-pulse image have almost the same height values resulting into a small value in the vegetation index image. Thus these trees are eliminated.

The FOM segmentation procedure based on region growing was proposed by Geibel and Stilla (2000). In each iteration, the two adjacent segments that are most similar to each other are merged. Similarity is expressed by the distance between two segments. The result has been evaluated in comparison to the following four segmentation methods:
Burns et al. (1986), Jiang and Bunke (1994) combined with Hoover et al. (1996), Flynn and Jain (1988) combined with Hoffman and Jain (1987) and Hoover et al. (1996). According to a quality measure proposed by Geibel and Stilla (2000) the FOM segmentations are of higher quality than the other segmentations. The quality is measured based on the amount of over- and under-segmentation.

Multiple thresholding is employed by Zhan et al. (2002) for the segmentation of laser data. Threshold values are selected from 1 m to a maximum value producing a number of binary images. The regions generated from each thresholding step are evaluated using two main criteria namely size and location. The size is calculated as the actual number of pixels of the segment and the location is measured by the center of mass of the region. If the region has a little change in size and only a small shift of the center of mass between adjacent layers is classified as building. Other information sources are used to discriminate between building and vegetation regions such as spectral information or first pulse minus last pulse data.

In (Rottensteiner and Briese, 2002) DTM and nDSM are generated using the robust interpolation method proposed in (Kraus and Pfeifer, 1998). The initial building segments are produced by thresholding pixels higher than $3.5m$ and connected component labeling. After morphological refinement texture analysis is employed for separating vegetation from the initial building segments. For this purpose point features proposed by (Fuchs, 1998) are used. The first derivative of the DSM is calculated using a 9×9 kernel. For each initial building segment the number of "point-like" pixels is counted. Regions having more than 50% "point-like" pixels are classified as vegetation. Vegetation connected to buildings cannot be properly classified. Therefore, those regions are separated using morphological processing and evaluated again.

Vögtle and Steinle (2003) employ fuzzy logic when classifying laser pixels into buildings, vegetation, and terrain. At the beginning an approach named as " convex-concave hull" (von Hansen and Voegtle, 1999) is employed for generation of DTM from laser data. The normalized DSM is generated by subtracting the DTM from laser data which is used for segmentation of 3D objects. The gradients on the segment border, difference of first- and last-pulse height, shape, and height textures are the input data for the classification. The gradient values on the boundary of the segment are used to separate non-ground objects, i.e., buildings and vegetation, from the ground pixels. The difference between first- and last-pulse is employed to discriminate buildings and vegetation. The shape parameter is defined by "parallelism of long segment contour lines" evaluated by the deviation of line directions assuming that buildings borders are relatively long and parallel to each other.

The height texture is provided by the Laplace operator, as employed by Maas (1999b), and

local curvature (Steinle and Vögtle, 2001). The former two criteria, i.e., shape and height texture, assist the difference of last- and first-pulse for separating buildings and vegetation.

Vosselman et al. (2004) proposed to use surface roughness, segment sizes, as well as shape and color (if available) in addition to the vegetation index defined by the difference between first- and last-pulse data for the classification of ALS data. They reported that 85% of the building points and 78% of the vegetation points were classified correctly.

A hierarchical procedure for the segmentation and classification of buildings and trees has been proposed by Arefi and Hahn (2005). The starting point for the hierarchical process are morphological operations with different SE sizes applied to last pulse LIDAR data. The key of the segmentation process is the analysis of the generated sequence of morphologically filtered data to extract ground points with high probability. Aggregation to regions and the extraction of region properties provide the basis for 3D object extraction.

Hebel and Stilla (2008) proposed a method to detect the ground level and to separate clutter and man made objects in airborne LIDAR data. It is based on filtering the laser points in each scan line keeping the points that are similar to building parts. Points on the ground level and points belonging to irregularly shaped objects such as trees and vegetation are eliminated. RANSAC-based straight line detection is employed to distinguish between clutter and man-made objects. The straight lines are merged and modeled by polygons. In a related paper (Hebel and Stilla, 2007) it is shown now gaps due to missing structures of facades in airborne or terrestrial laser data can be filled by additional data captured by an oblique looking laser scanner. Multi aspects laser data has been acquired and then registered using an Iterative Closest Point (ICP) based algorithm (Besl and McKay, 1992).

Finally, some methods use other resources in combination with LIDAR data in order to separate buildings and vegetation data, for example, 2D GIS data (Stilla and Jurkiewicz, 1999; Brenner, 2000b), aerial images (Ameri, 2000), or multi-spectral imagery (Haala and Brenner, 1999).

3.1.3 3D Building Models

The determination of building parameters for their 3D reconstruction is a challenging problem addressed by many researchers. Ever since airborne LIDAR data appeared as a new data source in remote sensing and photogrammetry attempts have been made to model buildings using LIDAR data. LIDAR combined with aerial images was, e.g., used for building reconstruction by (Haala and Anders, 1997; Rottensteiner and Briese, 2002). The LIDAR data is employed to segment planar faces while the aerial image is used to improve the quality of the edges of the segments.

The first approach which used only LIDAR data for building reconstruction was presented by Weidner and Förstner (1995). They used two different 3D models: a simple parametric model for buildings with rectangle ground plans and a prismatic model for complex buildings.

Another model-driven approach was employed by Maas (1999a). He generated building models based on the analysis of invariant moments of segmented regions. He assumed that buildings consist of a limited number of structures such as gable roofs. Using the first and the second order invariant moment, a number parameters of a building such as position, orientation, length, width, height, and roof inclination are determined. The moments of all segmented regions are measured, and the building parameters are derived from these moments.

Building reconstruction integrating maps for the extracting of the outlines was proposed by Stilla and Jurkiewicz (1999). A prismatic model is generated employing the height values

inside the polygon. The roof structures are classified by histogram analysis of the points inside the polygon into flat roof, flat roof with superstructure, gabled roof, and hip roof. The gradient image is used to separate gabled from flat roofs. Faces are segmented based on the similarity in gradient values. When intersecting neighboring segments special attention is required to provide the appropriate common line.

Prismatic buildings are extracted by means of edge detection in (Alharthy and Bethel, 2002). The algorithm is devised only for buildings with rectangular shape.

Segmentation based approaches are proposed by Geibel and Stilla (2000) and Rottensteiner and Jansa (2002) to find planar regions which determine a polyhedral model. Geibel and Stilla (2000) used in their approach only one parameter to control the segmentation and showed a strategy to assess the segmentation results.

Another segmentation based approach using a TIN structure for the data is proposed by (Gorte, 2002): Segments are created by iteratively merging triangles based on similarity measures. Finally, the segmented TIN structures is transformed into a VRML model for visualization.

(Rottensteiner, 2006) described a model for the consistent estimation of building parameters which is part of 3D building reconstruction. Geometric regularities were included as soft constraints in the adjustment of the model. Robust estimation can be used to eliminate false hypotheses about geometric regularities.

A comparison between data- and model-driven approaches for building reconstruction has been made by Tarsha Kurdi et al. (2007). It states that the model-driven approach is faster and does not visually deform the building model. In contrast, the data-driven approach tends to model each building detail to obtain the nearest polyhedral model, but it usually visually deforms the real shape of the building. Since the modeling accuracy is strongly related to the technique used for building modeling, a meaningful comparison about the modeling precision of these two approaches is not possible.

3.2 3D Building Reconstruction Using Other Data Sources

In this section building reconstruction employing other data sources with or without integrating LIDAR data is reviewed. It contains two parts: In the first part algorithms while only use image data for building detection and reconstruction are summarized which in the second part methods which integrate LIDAR data with other data sources are explained.

3.2.1 3D Building Reconstruction from Imagery

Photogrammetry is used to collect spatial data such as buildings from aerial photographs which are available around the world. Stereo pairs are used to extract the 3D objects located in the overlapping areas of the photographs. Orientation parameters are the information required to find the mathematical relationship between the photographs and object space. This basic relationship between image and object coordinate system can be found in all photogrammetric literature such as (Mikhail et al., 2001) and (McGlone et al., 2004).
Currently, several photogrammetric processes are utilized mostly automatically. These comprise for example, aerial triangulation and production of DTM with digital image correlation

(Baltsavias et al., 2001). In the past few years, the increasing number of applications for 3D city models, such as, urban planning, environmental analysis, and tourism, led to a continuous increase of the demand for 3D data. Due to the costly measurement of 3D objects using stereo photogrammetry, there was a constant development of automated 3D acquisition methods.

The development of cost-effective systems aims at simple hardware and relatively untrained operators. For automatic 3D object reconstruction it is necessary to derive complete object structures including the object topology from measured points. Fully automatic acquisition of 3D buildings is not yet possible and in practice, only semi-automatic systems can be found (Gülch et al., 2004). Nevertheless, these systems are advantageous compared to a fully manual measurement.

Building reconstruction from image data comprises two main parts, namely, building detection and building reconstruction. There is plenty of research using image data, mostly aerial images. Reviews can be found in (Mayer, 1999; Gülch et al., 2004). Often, stereo images taken from aerial or satellite photography are used to generate a DSM. From it a normalized DSM (nDSM) is computed employing morphological operations which is the basis for segmentation based on thresholding. The produced DSM and nDSM are in most cases of lower quality than from LIDAR data. Brenner (2000a) compared the quality of a DSM generated by image correlation using the Match-T software (Ackerman and Krzystek, 1991) with LIDAR DSM. The comparison of the contour lines produced from the two data sets showed that the quality of the LIDAR data, especially on building edges, is higher.
Weidner and Förstner (1995) employed the above steps to produce DSM and nDSM. The non-ground regions are classified into buildings, vegetation, and other classes after segmentation of the nDSM taking the size and position of the building bounding boxes as criteria. The minimum size of a segment is related to the expected minimum size of a building. Additionally, the position of the bounding box is used to filter the segments which are not entirely inside the image. Parametric models are used for modeling simple and detached buildings using few parameters, i.e., length, width, and height of the building. Complex buildings and building blocks are described using prismatic models. For approximation and regularization of the building outline an MDL based (Minimum Description Length) approach is employed. It fits a rectangular polygon to the building outline. After estimating an average height, the prismatic model is formed.

Nevatia et al. (1997) described a method for detecting and constructing rectilinear buildings with flat roofs from single aerial images. They use geometric and projective constraints to construct hypotheses for the presence of building roofs from linear features in the image. Shadow and wall evidence are used to verify and reconstruct 3D buildings. Results from several images can be integrated into one model.

Baillard et al. (1999) presented an approach for the automatic 3D reconstruction of buildings from multiple images. Line features are extracted in all images and matched by the method proposed in (Schmid and Zisserman, 1997). Epipolar geometry is used to provide point to point correspondences on putatively matched line segments from two images. The similarity of the lines' neighborhoods is assessed by cross-correlation at the corresponding points. The matched lines lead to a piecewise planar reconstruction of the building.

Ameri (2000) segments Region of Interests (ROIs) for individual buildings by the method proposed in (Brenner, 2000a). By least square estimation, planar regions and roof surfaces are recognized. The topological relations between the roof elements are defined by means of Voronoi Diagrams leading to a rough building model.

Elaksher et al. (2002) employed multi image line matching to extract 3D building wire-frames. They used multiple images to deal with the occlusion problem often existing in stereo processing. Homogeneous regions are extracted from each image using the split-and-merge algorithm. Classification of the regions is carried out based on size, shape, and intensity values to extract roof regions. The border lines for the building regions are extracted using the Hough Transform. Next, roof segments are matched pair-wise among multiple images using the algorithm described by Scott and Longuet-Higgins (1991).

Object extraction based on neuro-fuzzy modeling was proposed by Samadzadegan et al. (2005). 3D regions are extracted from the DSM using morphological processing and for 2D regions a fuzzy-based region growing technique is employed. Structural, textural, and spectral (STS) information is extracted and integrated by fuzzy reasoning. A neural network supports the fuzzy logic approach for performing automatic image classification.
A processing chain for generating automatic 3D city models from very high resolution (VHR) stereo satellite images was presented by Kraus et al. (2007). A DSM is produced using image matching from the stereo IKONOS images. Using an adapted median filter a DTM is computed from the DSM. After calculating the nDSM from the DSM and the DTM, a high resolution true orthophoto is generated using the DSM, the Rational Polynomial Coefficient (RPC) parameters, and pansharpened multispectral stereo images. The true orthophoto, NDVI, and the nDSM are employed to classify the pixels in stereo images. With a model based technique 3D models of buildings, trees, and ground regions are reconstructed. For buildings a coarse prismatic model is generated in the form of a polygon and a flat top with the average height extracted from the DSM.

Besides the classical photogrammetric approaches to determine object heights indirectly, range sensors like LIDAR and interferometric SAR (InSAR) play an important role since recent years. LIDAR and InSAR are both active systems which illuminate the scene with electromagnetic waves and measure the backscattered signal component. Due to the different wavelengths (e.g. LIDAR: 1.5 mm, SAR-X: 3cm) special surface properties are sensed, because the reflectance depends on the surface roughness compared to the wavelength. Stilla et al. (2001) compare in their contribution segmentations of LIDAR and InSAR data for building reconstruction using data having the same ground sampling distances (GSD=1m). It was shown that LIDAR allows a much more detailed reconstruction of geometric building details than InSAR data with comparable GSD. An inherent drawback of InSAR for building reconstruction is the side-looking geometry (Stilla et al., 2003). Furthermore, Stilla (2007) shows that SAR specific effects like layover, shadow, multiple scattering cannot be neglected and are difficult to analyze for building reconstruction from SAR data of very high resolution (GSD=0.1m)

Mayer (2008) discusses the current state and promising directions of automatic object extraction in photogrammetric computer vision. He notes that there are only few systems available in the market which successfully work in practice. A number of issues which are important to be considered with this respect are mentioned. For practical applications the following reliable and robust techniques were recommended:

- Scale Invariant Feature Transform (SIFT) operator (Lowe, 2004) for robust scale- and rotation-invariant point extraction and matching. In addition, they are also robust to change in illumination, noise, and minor changes in viewpoint (Wikipedia, 2009).
- 5-point algorithm (Nister, 2004) to directly compute approximate values for calibrated relative orientation.
- RANSAC for robust estimation suitable even for well below 50% correct data.

3.2.2 3D Building Detection and Reconstruction by Combination of LIDAR and Other Data Sources

Baltsavias (1999) comprehensively compared airborne laser scanning and photogrammetry. Advantages and disadvantages of each was explained particularly concerning data quality and it was concluded that LIDAR and photogrammetric data are complementary to each other.
Haala and Brenner (1999) employed multi-spectral and LIDAR data to classify building and vegetation regions. They used LIDAR together with building ground plan data to generate 3D building models.

The ground plan induces strongly the shape of the roof. In contrast to this Stilla and Jurkiewicz (1999) proposed an approach which use LIDAR data and a ground plan in a similar way, but allows to reconstruct roof structures and superstructures which could not be predicted from the outline of the building.

For the combined exploitation of LIDAR data and existing ground plans for building reconstruction Vosselman and Dijkman (2001) employed two strategies: The first one is based on the detection of intersection lines and height jump edges between planar faces. In the second strategy, a coarse 3D model is refined by analyzing the points that do not fit well.

A feature-based fusion of LIDAR data and stereo imagery is presented by Schenk and Csatho (2002). Features such as straight lines and surface patches are detected in each data set and correspondences between them are established. After providing a common reference frame, image fusion of geometric and semantic information leads to an explicit surface description.
McIntosh and Krupnik (2002) tried to improve the quality of the DSM derived from LIDAR data using information extracted from image data. 3D line segments are derived from stereo images and then registered with the LIDAR DSM.
Hu et al. (2004) improve the quality of building models generated from LIDAR data by integrating aerial imagery. A "primitive-based" model refinement is employed based on the shape of the building roofs. The are classified into two groups: linear-fitting primitives and higher-order surface primitives. After the segmentation of the building roofs a geometric primitive is fitted to each. Image textures and color information are used to refine the building model.
Rottensteiner et al. (2004a,b) proposed a method using the data fusion theory of (Klein, 1999) based on "Dempster-Shafer" combination of evidence for classification of LIDAR data and multi-spectral images. Additionally, aerial images and LIDAR data were combined to improve the results of roof-plane detection and roof shape boundary delineation (Rottensteiner et al., 2004a).

The gap of missing structures of facades in classical nadir airborne laser data and missing structures from roofs in terrestrial laser data can be closed by capturing the scene with an oblique looking laser scanner. Hebel and Stilla (2007) show a multi aspect full waveform acquisition, registration, and fusion of point clouds for capturing the roofs and all facades of buildings. Full waveform LIDAR provides an intensity value for each laser point and allows a direct texturing of the scene. Due to the possibility to model roof and fasade by faces makes it much more easier to seprate buildings from other structures.

Chapter 4

A Novel Approach for Filtering of LIDAR Range Data Based on Geodesic Morphology

In this chapter, the proposed methods for the classification of the LIDAR pixels into ground and non-ground regions are discussed. A hierarchical segmentation procedure using geodesic image reconstruction is described following (Arefi and Hahn, 2005) to separate 3D non-ground points from ground points. The key of the segmentation process is the analysis of a generated sequence of morphologically filtered data to extract ground points with high probability and separate them from non-ground points. The procedure continues by the interpolation of the ground points to generate a Digital Terrain Model (DTM). Finally, the DTM is subtracted from the original range image resulting in the Normalized Digital Surface Model (nDSM). By this subtraction pixels corresponding to uneven ground surfaces are suppressed and, therefore, the non-ground regions can be thought to be located on a horizontal plane with *zero* height.

4.1 Pre-processing

Before filtering with geodesic morphological reconstruction outlier removal and data rasterization are inevitable. The following two sections deal with these two issues.

4.1.1 Outlier Detection and Removal

This section explains the automatic detection and elimination of outliers from the point clouds. An outlier is an observation not consistent with other observations, particularly in its neighborhood. In the given case, in large datasets outliers are points which are far away from the sample mean in the local neighborhood. Figure 4.1 illustrates a side view of 3D laser points around a building block. The dataset contains several outliers which are all below the ground.

According to the ISPRS filtering report (Sithole and Vosselman, 2003), two different types of outlier often exist in the LIDAR data sets, namely *low points* and *high points*

- Low points: They are located lower than other points in a neighborhood area. Low points do not belong to the topography and usually originate from multi-path reflection errors and errors in the laser range finder (cf. Figure 4.2).

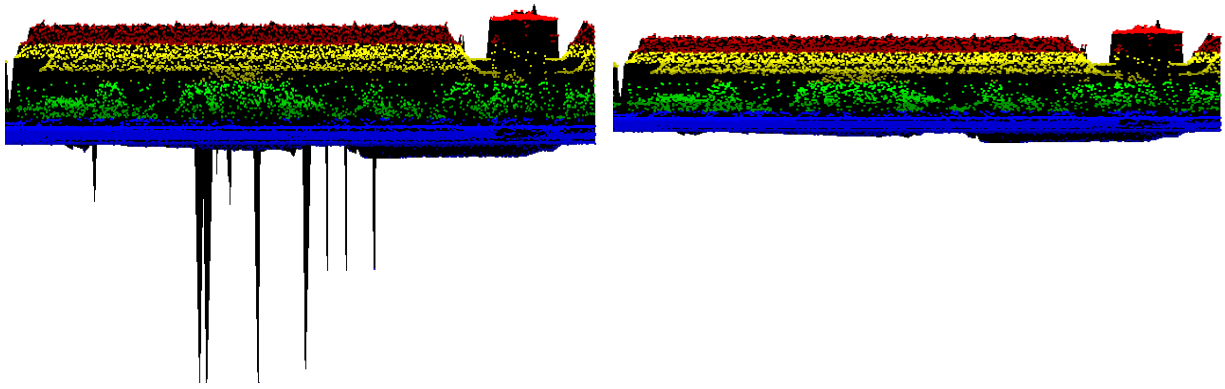


Figure 4.1: Laser data with outliers (left) and after outlier removal (right)

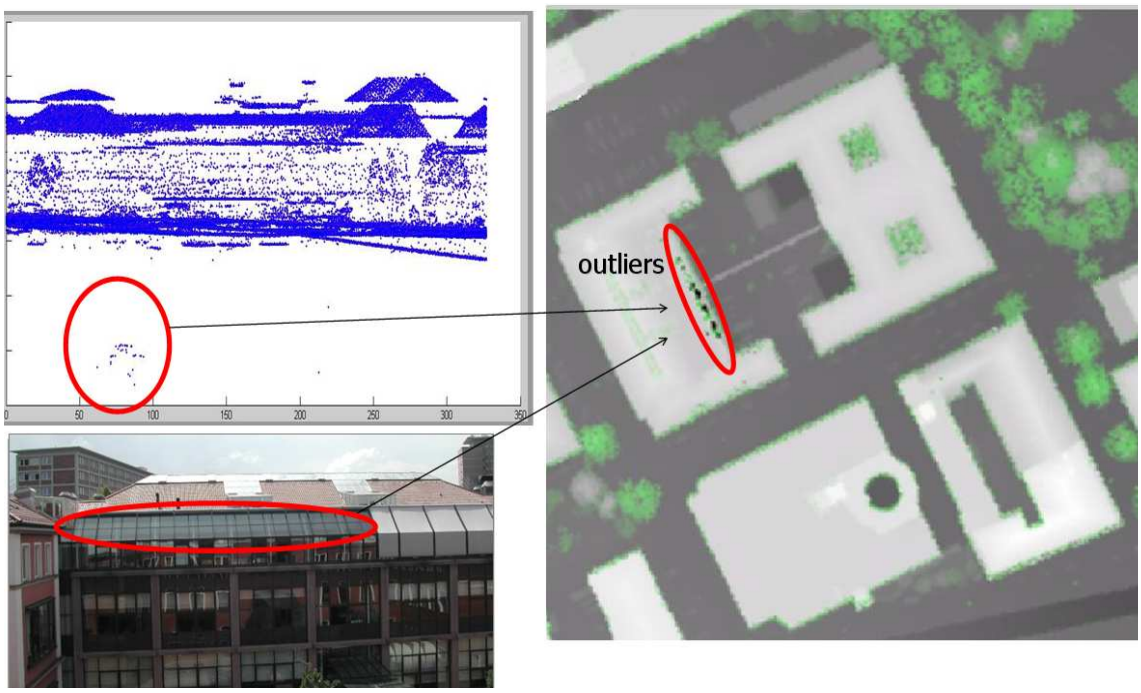


Figure 4.2: Glass structure on building results in multi-path reflections (outliers)

- High points: They originate from laser pulses reflected by objects below the aircraft, such as birds. They are located at an elevation obviously higher than other points in their neighborhoods.

There are many statistical methods available to distinguish the outliers and eliminate them from the point clouds. Some methods assume, that the data stems from a normal distribution and therefore, they identify the observations which are deemed "unlikely" based on mean and standard deviation (Wikipedia, 2008). Two statistical procedures which are widely used in practice to detect and remove outliers are:

1. "Grubbs' Test" (Grubbs, 1969): For a given dataset, the minimum and maximum values are calculated and the following values are computed:

$$G_{min} = \frac{\bar{X} - X_{min}}{s}$$

$$G_{max} = \frac{X_{max} - \bar{X}}{s}$$

where $\bar{X}$, s, X_{min} and X_{max} denote sample mean, standard deviation, minimum, and maximum value, respectively. For the two-sided test, the hypothesis of no outliers is rejected if:

$$G > \frac{(N-1)}{\sqrt{N}} \sqrt{\frac{t^2_{(\alpha/(2N),N-2)}}{N-2+t^2_{(\alpha/(2N),N-2)}}}$$

where $t_{(\alpha/(2N),N-2)}$ denotes the critical value of the t-distribution with $(N-1)$ degrees of freedom and a significance level of $\alpha/(2N)$ (NIST/SEMATECH, 2007).

2. "Inter Quartile Range *(IQR)*" (Martinez, 2004): It is a measure of statistical dispersion and it is equal to the difference between the lower $q(0.25)$ and the upper $q(0.75)$ quartile of the dataset. (The lower quartile is the value for which 25% of the data are less than or equal.)

$$IQR = q(0.75) - q(0.25) \tag{4.1}$$

Two parameters are determined based on the IQR to limit the range of accepted data or to detect outliers. These two parameters are the "lower limit" LL and the "upper limit" UL and are determined as follows:

$$LL = q(0.25) - 1.5 \times IQR \tag{4.2}$$
$$UL = q(0.75) + 1.5 \times IQR \tag{4.3}$$

Then observations smaller than LL and larger than UL are defined as outliers with respect to the bulk of the data.
"Box-and-whisker diagram" or "Box-plots" a way to visualize summary of the statistics.

Figure 4.3 shows an example. The inliers limited by UL and LL are illustrated by the black range limited by the symbols $\top$ and $\perp$, respectively. The lower quartile $Q1$ and upper quartile $Q3$ are shown as well Extreme outliers far away from the median with the Extreme upper limit $Q3 + 3 * IQR$ and the Extreme lower limit $Q1 - 3 * IQR$.

3. Terrasolid (2007) method: In this thesis the routine in the *Terrascan* software (Terrasolid, 2007) is used to detect and remove low (negative outliers) as well as high points from the LIDAR data. Two different routines are used to detect these two types of outliers.

   (a) Negative outliers: For each point a neighborhood is considered and the low points are the points with a height value less than a pre-defined threshold below all other points within a given xy distance. However, this routine can also search for groups of low points where the whole group is lower than other points in the vicinity

   (b) High points: Points which are higher than a pre-defined value called *Limit* times the standard deviation above the median elevation of the surrounding points are classified. As for the negative outliers, a given search radius is employed.

- **Comparison of Outlier Detection Techniques:** Three methods have been explained to detect and eliminate outliers from LIDAR point clouds. The first two are common statistical algorithms which are widely used to extract outliers from normally distributed sample data. The assumed normal distribution might be correct in open terrain, but in

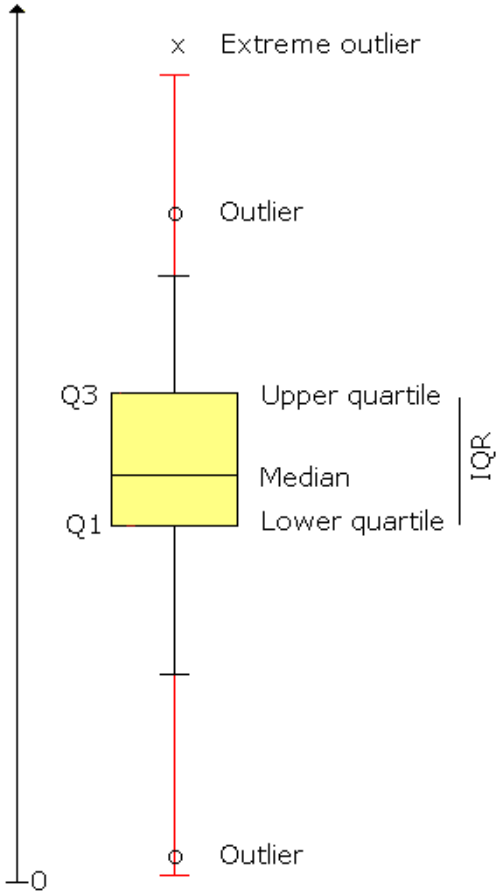


Figure 4.3: An example for data visualization using Box-plot

other regions such as especially urban areas, where many buildings are located, it is not appropriate (ASPRS, 2004).

Grubbs' tests require more computations than the IQR based method, but they are more powerful if the assumption of normal distribution is fulfilled. They can be properly used for outlier removal from DTM as well as digital elevation models (DEM) with wider point spacing, e.g., $10m$. For the latter, the data set in general, but also smaller window tiles, e.g., $200m \times 200m$, follow normal distribution.

In contrary, the Terrasolid method locally evaluates each suspicious point checking whether the potential outlier point is clearly above or below the other points in its adjacency. For this method the data set does not need to follow normal distribution and it is thus suitable for the data employed in this thesis.

4.1.2 Data Rasterization

Data rasterization is the process of converting vector data to a raster image. Since reconstruction by means of geodesic dilation is much more easily and efficiently formulated as an image processing technique applied to gridded data, the generation of an image using spatial interpolation methods is a must. On the other hand, interpolating the data to fixed point intervals enables the use of kernel filtering functions. As result of interpolation, the row and column position of a given LIDAR point in a DSM matrix corresponds to its longitude and latitude,

respectively. The value of the cell in the matrix corresponds to the surface elevation at the interpolated location.

Many techniques are available for spatial interpolation such as kriging, inverse distance weighting (IDW), nearest neighbor (NN), minimum curvature, polynomial regression, and moving average. Some of these algorithms work properly for some application and fail in others. Most of the interpolation routines predict the unknown points using the values of the closest neighbors by fitting a model to them.

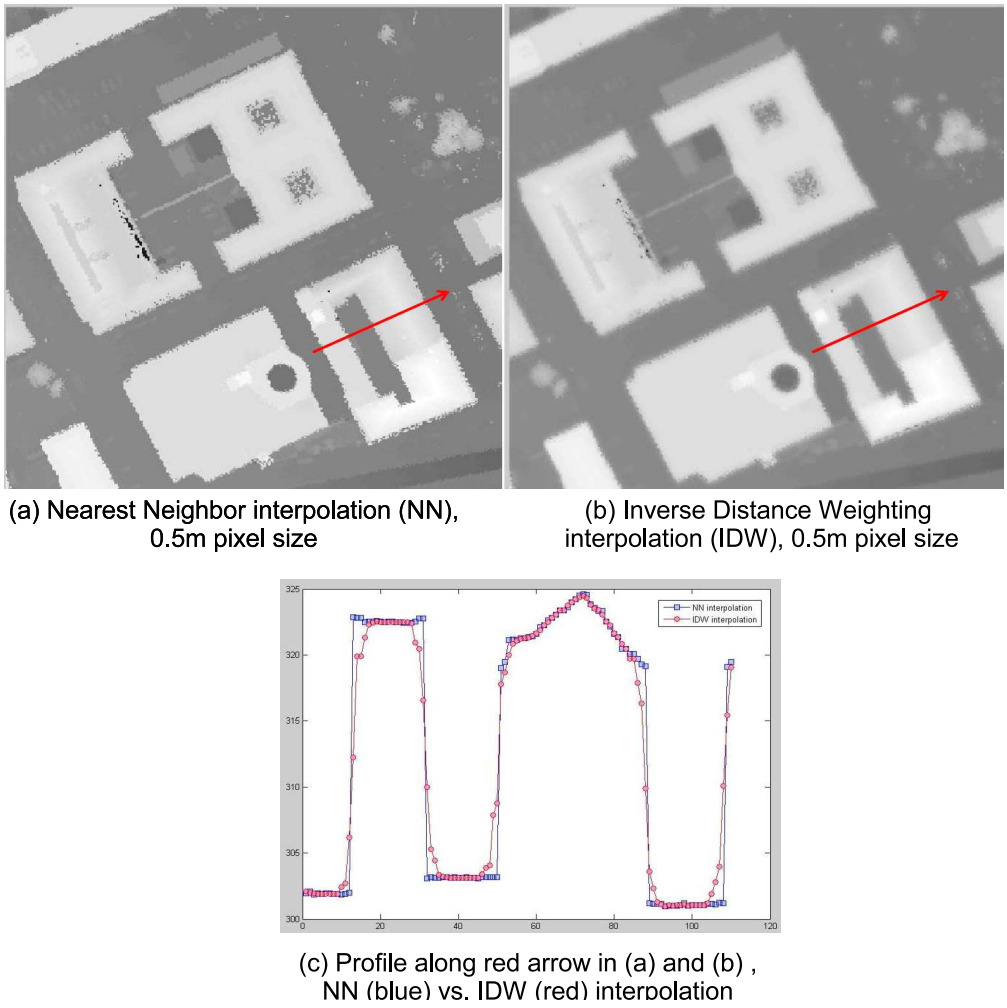


Figure 4.4: Comparison of Nearest Neighbor (NN) and Inverse Distance Weighting interpolations by evaluating building edge pixels using a profile along the red arrow

The ultimate goal of filtering the laser data is to separate non-ground pixels from the ground pixels and generate a DTM or nDSM. A DTM is a topographic model of the "bare earth" that enables users to infer terrain characteristics that may be hidden in the DSM. It has had vegetation, buildings, bridges, and other cultural features removed, leaving just the underlying terrain (INTERMAP Technologies, 2008).
Therefore, for any 3D object such as a building only the lowest part and not the points on its body, e.g., on the walls of the building are of interest. Thus, an interpolation approach

which does not smooth points located on the edges of the 3D objects seems to be most suitable. Nearest neighbor interpolation simply choosing the nearest gray value or height regarding to a grid cell is such an approach. In this section two interpolation techniques are explained and their results especially on building edges are discussed. The first method is nearest neighbor (NN), which simply selects the value of the nearest point, and does not consider the values of other neighboring points at all. The second interpolation method is IDW, which basically assumes that closer points to the interpolated point have more effect (weight) than the other points. This technique creates weights according to the distances between the interpolated location (x, y) and each of its neighbors. In other words, the weight functions should be largest at *zero* distance and decrease as the distance increases. The weight functions or "power functions" are provided by $y = x^{-P}$ with $P > 0$ which decrease as x increase. The following steps are involved in IDW interpolation:

- Selecting points $(x_1, y_1), (x_2, y_3), ..., (x_n, y_n)$ in the neighborhood of (x, y).
- Compute the weight $W(i)$ values for each point (x_i, y_i) in the neighborhood as:

$$W(i) = \frac{1}{d_i^P} = [distance[(x, y), (x_i, y_i)]]^{-P} = [(x - x_i)^2 + (y - y_i)^2]^{-P/2}$$

- Normalize the weights by dividing them by the *sum* of all the weights and compute the interpolation value z as:

$$z = \frac{\sum_{i=1}^{n} \frac{z_i}{d_i^P}}{\sum_{i=1}^{n} \frac{1}{d_i^P}}$$

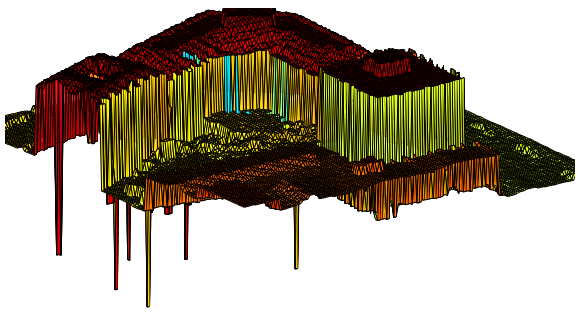
  where z_i is original value of the point (x_i, y_i).

Figure 4.4 displays interpolation results of a sample point cloud by means of NN (cf. Figure 4.4(a)) and IDW (cf. Figure 4.4(b)) techniques. A visual interpretation shows that the image provided by IDW is blurred and it is smoother than NN. A profile generated along a building verifies this conclusion, as it is illustrated in Figure 4.4(c). The height values along the red arrow in Figures 4.4(a) and 4.4(b) are collected from both interpolated images and represented in Figure 4.4(c). The profile proves that the NN interpolation method preserves the height jumps of the building edges in a more appropriate way than IDW. Therefore, in NN interpolation all the pixels located at the building edges will almost have same height values with a certain height jump compared to the pixels next to them on the ground. On the other hand it is fair to mention that the accuracy of the NN interpolation can be worse than the linear interpolation, for instance on strongly inclined planar roof faces, if the point density is low.

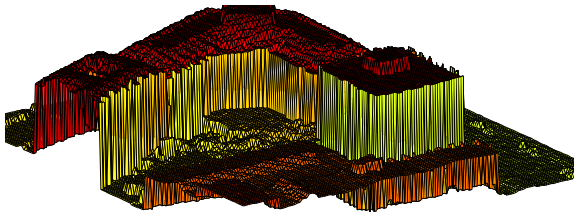
This property is very important for image reconstruction using geodesic dilation. The profile belonging to the building part with the gable roof (cf. Figure 4.4(c) middle) shows, that the building edge is located almost at a height of 320 and the pixel next to the building on the ground has almost a height of 303. Considering the building top height as 325, a geodesic reconstruction using a single offset h which is a value between $h_1 = 325 - 320 = 5$ and $h_2 = 325 - 303 = 22$ can properly filter the building part (cf. Section 4.2 and Figure 4.9 bottom). I.e., there is a high flexibility to select h as a value between 5 and 22 and this is not the case for the IDW image in which h should be carefully selected, e.g., about $h_2 = 325 - 303 = 22$.

Figure 4.5 represents an image generated using NN interpolation. As mentioned before, outliers occur in the point cloud as single points with high elevation difference compared to their

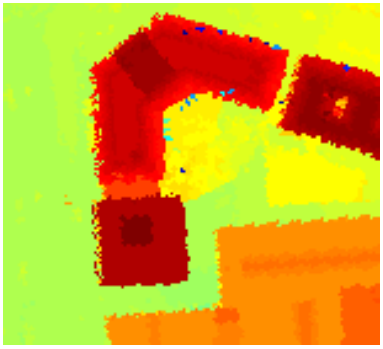
adjacent points. In regions where outliers are located in a gap, e.g., inside a water body, the single outlier points change into a region after NN interpolation and thus gaps might remain in the image. If the interpolation parameters are properly tuned, big gaps are mainly produced



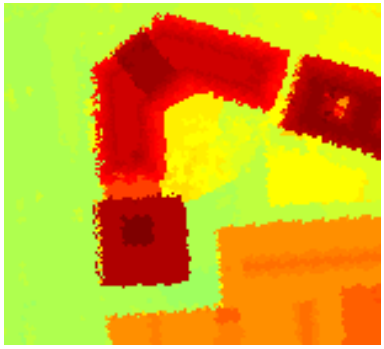
(a) Effects of outliers on rasterized data - 3D view



(b) Image data with outliers removed - 3D view



(c) Effects of outliers on rasterized data



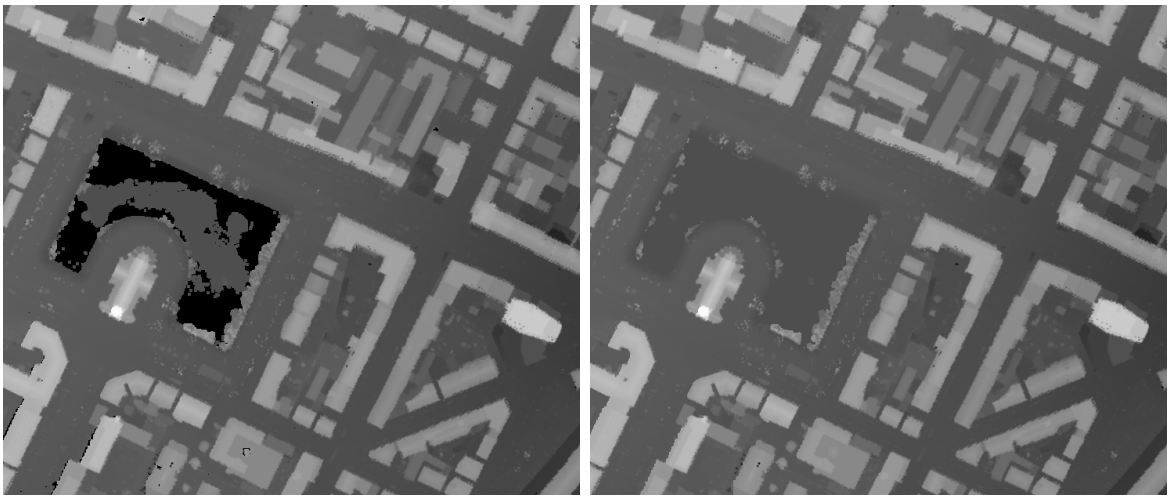
(d) Image data with outliers removed

Figure 4.5: Image generation by means of Nearest Neighbor interpolation

by very weak reflections, too weak for the receiver, or by mirrored reflections away from the receiver. The latter occurs for water bodies as well as glass structures such as glass roofs. The gaps should be filled by appropriate height values before applying geodesic morphology. A simple method to fill gaps works as following:

1. Pixels belonging to gaps are extracted. These regions usually have NAN values (Not-A-Number) or any other pre-defined value such as -10000.
2. The regions (gaps) are labeled using connected component analysis.
3. The outer boundary for each region is found by subtracting its binary image from its dilated image.
4. For each labeled region the height values belonging to its outer boundary are taken from the original laser image.
5. The *median* value is calculated and used to fill the gap for each corresponding region.

Figure 4.6(a) shows an ancient church surrounded by a water body. For the largest part of the water body, the laser beam was mirrored away from the scanner and is thus represented by black color (NAN value). In Figure 4.6(b) the pixels with NAN values are replaced by gray values calculated by the above method.



(a) A Church surrounded by water body - black color shows pixels without height values

(b) The water region after replacing the pixels without height by appropriate values

Figure 4.6: Filling of gaps due to an existing water body based on the median of the gray values of the pixels located at the outer boundary of a gap region

4.2 Separation of Ground and Non-ground Based on Morphological Reconstruction

The classification of the LIDAR points is carried out in two steps: In the first step, the non-ground regions are separated from the ground regions and in the second step, the non-ground regions are classified into buildings, vegetation, and other classes. In this section the first step is explained and the proposed method is discussed.
Figure 4.7 illustrates the overall strategy to classify a LIDAR range image into 'ground' and 'non-ground'. As the flowchart shows, image segmentation is employed to group the potential non-ground regions. Image reconstruction by geodesic dilation (cf. Section 2.3.2) is the core operation utilized for image segmentation. The potential regions are then classified into non-ground and ground. A rule based classification separates regions that show a certain discontinuity at their boundary. The criterion or feature descriptor should allow to discriminate between the ground and other 3D regions which were wrongly grouped in the segmentation. I.e., it should highlight the boundary pixels of those regions which have a certain difference in height (jump) to the neighboring ground surface. Therefore, a gradient based feature which highlights large jumps is used as feature descriptor.
The outcome is, that all 3D objects such as buildings, vegetation, vehicles, trains will be classified as non-ground. The goal of the segmentation for which the new geodesic image reconstruction is employed is to simplify and/or change the representation of the image into something that is more meaningful and easier to analyze (Shapiro and Stockman, 2001). Here, the aim is to partition the laser image into multiple regions which cover the entire image. The main characteristic of the segmented regions is that they have a significant jump at the border compared to adjacent pixels.
The first step for segmentation by means of geodesic reconstruction is to provide Mask and Marker images. According to the definition, the Mask image is the reference image and it is the limit for propagation of the Marker image. Here, the original range image with its no-height pixels already replaced by an appropriate height value is chosen as *marker* image.
As explained in Section 2.3.2, the *marker* image has a significant effect on the results of geodesic

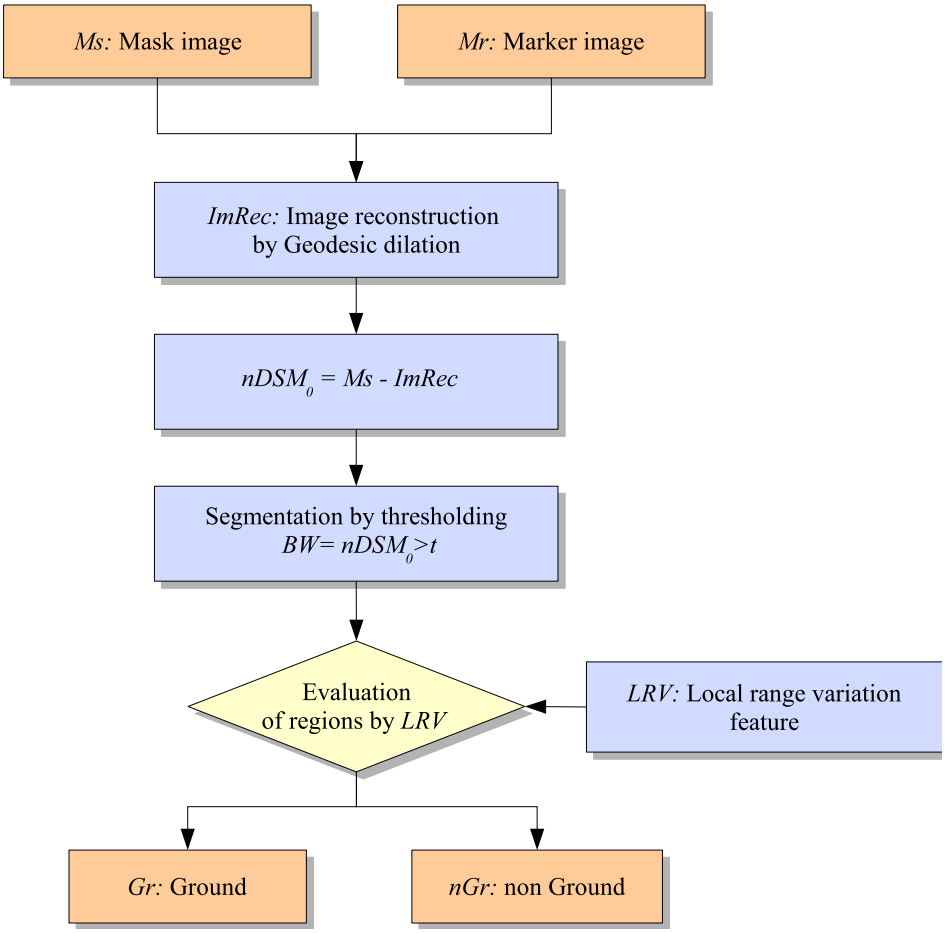


Figure 4.7: Separation of ground and non-ground

morphology. Some essential points have to be considered to create *marker* images which lead to suitable image segmentations:

- According to the definition, the *marker* image should have the same size as the mask image and the gray values should be equal or less than that of the *mask* image. In general, an appropriate *marker* image is determined using knowledge about the envisaged result or the physics of the objects it represents (JÄHNE et al., 1999a). Most commonly the *marker* image is generated by subtracting a constant value from the *mask* image as illustrated in Figure 4.8 (according to equation (4.4)). In this equation Mr, Ms, and h denote *marker* image, *mask* image, and $offset$ value.

$$Mr = Ms - h \tag{4.4}$$

  After choosing the offset value, the next step is to calculate a geodesic dilation of size 1 of the marker image with respect to the mask image based on Equation (2.13) . According to Equation (2.14), this process is continued until the pixel values do not change any more. The result of the successively performed geodesic dilations is the morphologically reconstructed image (cf. Figure 4.8 top right). An initial normalized $DSM(nDSM_0)$ is obtained by subtracting the reconstructed image from the original image (cf. Figure 4.8 bottom). In this figure the part which is represented by the reconstructed image is suppressed.

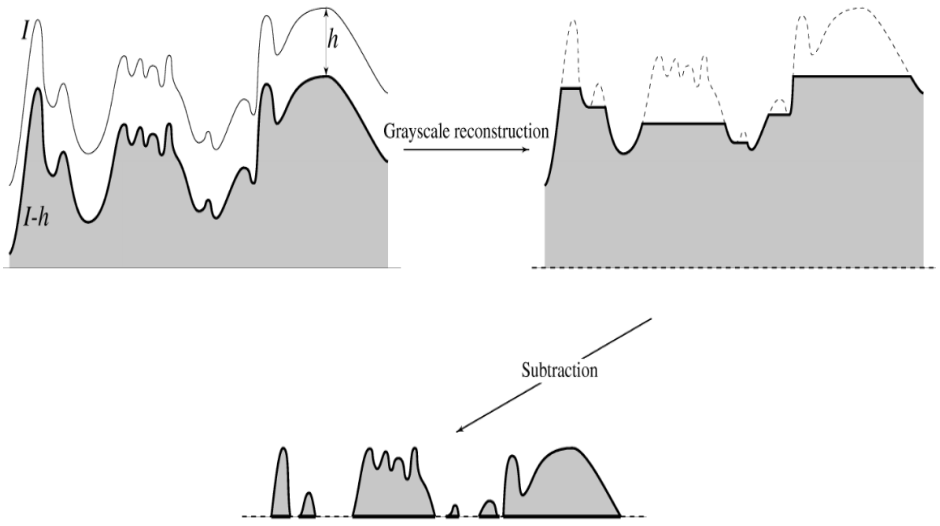


Figure 4.8: Reconstruction by geodesic dilation of a 1D mask signal I from a marker signal J. The marker signal is produced by subtraction of a constant offset h from the mask. The result of grayscale reconstruction using geodesic dilation is shown in the top right part. The subtraction of the reconstructed signal from the mask signal leads to the result given on the bottom (VINCENT, 1993)

- For the reconstruction, selecting an appropriate offset value h is critical, as it depends on the objects to be filtered. Figure 4.9 shows how the offset value should be chosen for filtering of a single building. The top row displays the intermediate result for a *marker* image with a small offset from the *mask* image. The result shows, that the resulting $nDSM_0$ is not acceptable, because the largest part of the building is filtered out and only a small part of the roof top remains after subtraction. In contrast to the first row, in the second row a big offset value is selected. The resulting marker image leads to an $nDSM_0$ which is also not acceptable: Not all pixels belonging to the ground are suppressed and are thus still in the $nDSM_0$. The third row demonstrates how a suitable selection of the marker image leads to a proper $nDSM_0$. Here h is chosen as the distance from the highest point on the roof to half the height of the building body. The produced $nDSM_0$ does not contain pixels belonging to the ground any more. Since the nearest neighbor interpolation is employed to transform the laser data into grid form avoiding intermediate heights at the boundaries of the buildings, this $nDSM_0$ is acceptable to extract the above the ground regions. For an ideal $nDSM$, the offset value has to be set to $h2$. In this case the building will retain its full height. If the goal of filtering is to enable the extraction of above the ground regions, an appropriate offset value should be between $h1$ and $h2$ as displayed in Figure 4.9 last row, left.
  We thus propose to use a sequence of offset values to create a sequence of marker images avoiding problems caused by an improperly selected offset. Formally, h is decreased from a maximum to a minimum value:

$$Mr = Ms - h \tag{4.5}$$

$$h = max(h) : -inc : min(h) \tag{4.6}$$

  Beginning of offset value with a bigger h (maximum) has this advantage that the regions with smaller h are all included in the regions detected with the bigger h. Therefore, e.g.,

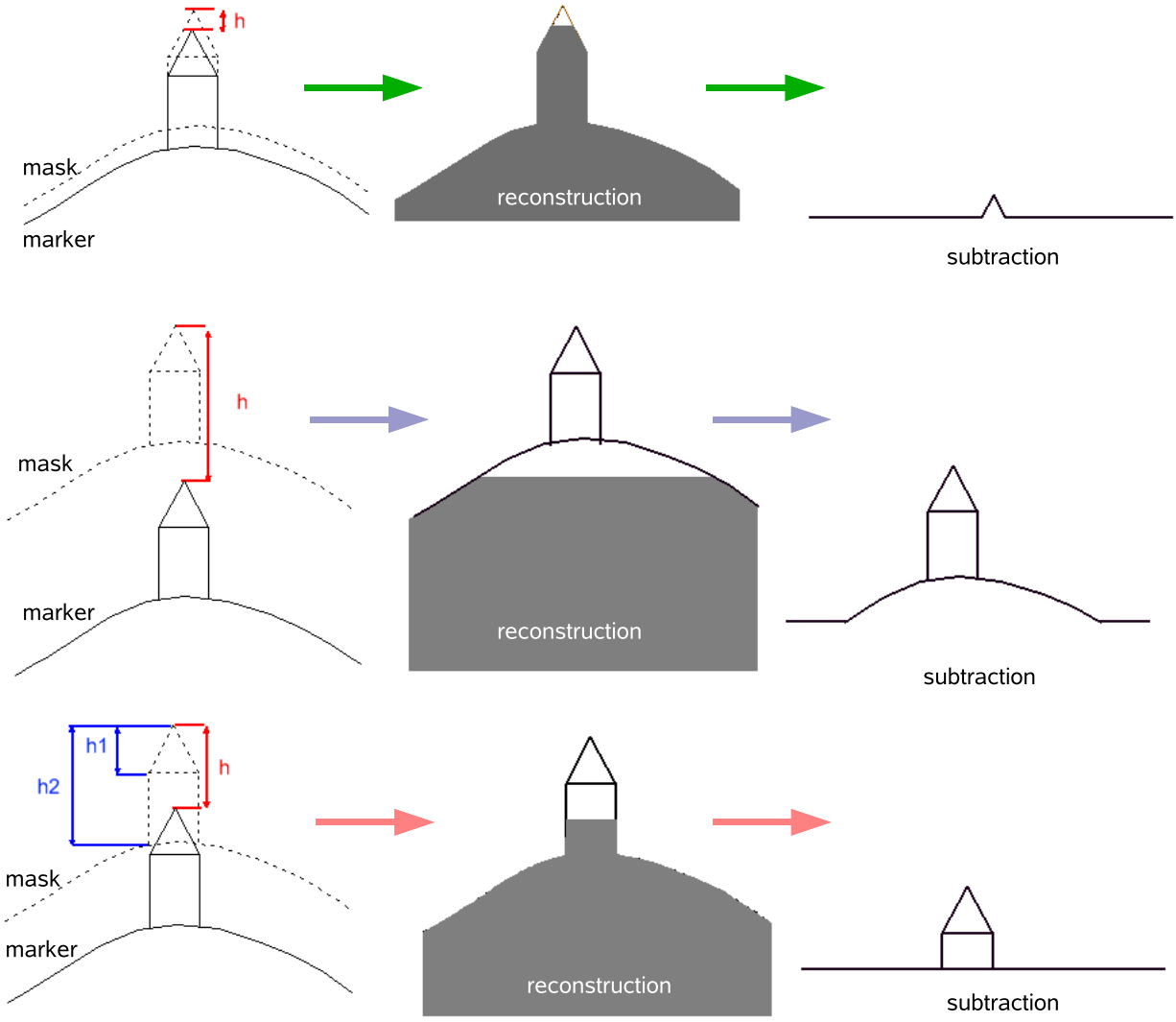
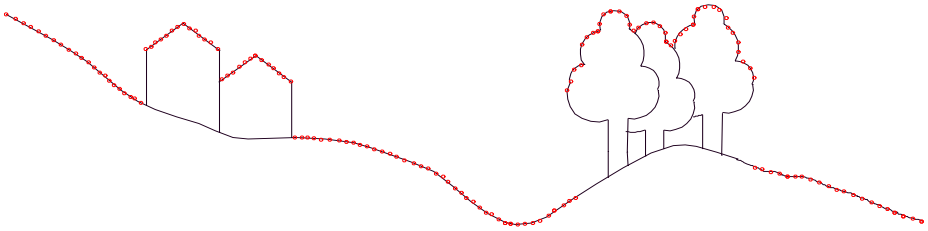


Figure 4.9: Selecting proper offset value to produce marker image

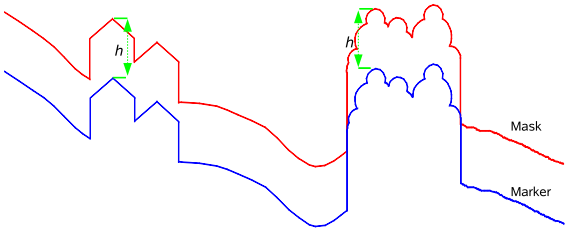
when a larger part of a building is detected and eliminated from the original image, in the next iterations the smaller parts will not appear any more for further evaluation. This makes process much more faster than process which begins with smaller h to a bigger one.

A sequence of *marker* images is produced by using an increment (*inc*). The *min* and *max* values are chosen according to the smallest and largest occurring distances from the top point of an object to the lowest ground point in its adjacency. To determine this parameter automatically, an image of the local height changes is calculated. This image is later used for the classification of the segmented regions into non-ground and ground based on subtracting the *maximum* from the *minimum* height change values in 3×3 local windows.

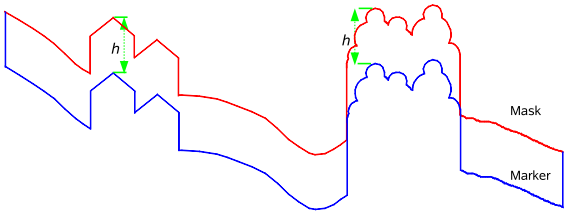
- Figure 4.10 represents another example for morphological geodesic dilation for a hilly terrain area with buildings and trees. The profile of the 3D scene is represented as black lines while the position of the collected laser points is highlighted by red dots (cf. Figure 4.10(a)). As shown in the left column, a *marker* image produced by a single offset value can succeed in filtering all $3D$ objects. The only weakness of the result is that the hilly region at the left side is not filtered. To filter regions connected to the boundary of the



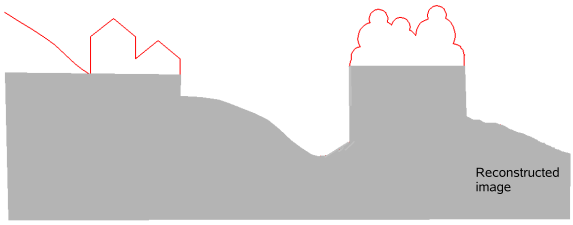
(a) Profile representing the ground (black color) and the location of the laser points (red dots)



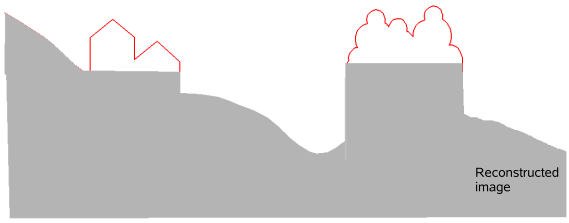
(b) *Mask* and *Marker*; *Marker* = *Mask* - *h*



(c) *Mask* and *Marker*; *Marker* = *Mask* - *h* except pixels at the boundary of the image when *Marker* = *Mask*



(d) Geodesic image reconstruction of (b)



(e) Geodesic image reconstruction of (c)

(f) $nDSM_0$ of (d)

(g) $nDSM_0$ of (e)

Figure 4.10: Geodesic image reconstruction by selecting a *marker* image by subtracting h as offset value from the *mask* image (left). In the right column the pixels at the image boundary are given the same values as the *mask*.

image, a small modification is proposed in form of not conducting the subtraction from the *mask* image at the boundary pixels of an image which thus remain the same as for the *mask* image (cf. 4.10(c)). This produces the result given in Figure 4.10(e) and 4.10(g).

- So far a proper image segmentation is obtained by choosing a sequence of *marker* images produced by different offset values from a maximum to a minimum value. Yet, since, for example in hilly residential areas, the height jumps for tall buildings can exceed one hundred meters, the offset value starting, e.g., at $min(h) = 1m$ should range to $max(h) = 100m$ making the whole procedure very time consuming.

  To speed up the process, image reconstruction begins with a *marker* image produced by an image with height values equal to the minimum height value of the *mask* except for the

image boundary pixels for which the height is left the same as for the *mask*. Depending on the surface undulation, a great number of objects will be filtered by selecting this *marker* image. Typically, between 60% to 80% of the objects are filtered. Thus, for the rest of the objects a fewer number of iterations is sufficient in the hierarchical process.

- in summary for the novel approach the following holds:

  1. The $Mask$ image is taken to be the original last pulse image after filling gaps caused by pixels with no height.
  2. $Marker$ images are provided by the sequential approach as follows:
     - The procedure begins with an initial *marker* image created by:

$$Mr(i,j) = \begin{cases} min(Ms(:)) & \text{for } (i \neq 1,m) \text{ and } (j \neq 1,n) \\ Ms(i,j) & \text{for } (i = 1,m) \text{ and } (j = 1,n) \end{cases}$$

       with $size(Ms) = [m,n]$, in which m is the number of rows and n the number of columns. I.e., the boundary pixels of the *marker* image take the values of the *mask* image.
     - The process continues by selecting sequential marker images produced by:

$$Mr(i,j) = \begin{cases} Ms - h, h = h_2 : -inc : h_1 & \text{for } (i \neq 1,m) \text{ and } (j \neq 1,n) \text{ and } (h_2 > h_1) \\ Ms(i,j) & \text{for } (i = 1,m) \text{ and } (j = 1,n) \end{cases}$$

       Since most of the objects are already filtered by the initial *marker* image, h_1 and h_2 are determined empirically in terms of percentages of the maximum and minimum height changes in the image. 10% of the maximum height change is selected for h_2 and $1m$ for h_1. Additionally, an increment value inc of $3m$ was determined empirically.

After carrying out the morphological image reconstruction based on the selected *mask* and *marker* images, an initial normalized DSM ($nDSM0$) is obtained by subtracting the reconstructed image from the *mask* image (cf. Figure 4.7). An initial classification of ground and non-ground points is carried out by binarising the $nDSM_0$. Any point in the $nDSM_0$ above a threshold (t) is classified as ground or non-ground point. These points, i.e., potential non-ground regions, are formed as regions by calculating connected components.

Theoretically selecting a zero value as threshold can be used to extract all the foreground objects but evaluating very low height objects such as small bushes has less importance in DTM production and make the process very time consuming therefore, in practice a value of, e.g., 30 cm, is selected to extract all significant above ground objects.

For the further analysis, features are determined for each region. Here, the size of a region and the local average height difference along the boundary of the region are used. By subtracting maximum and minimum values in local 3×3 windows along the region boundary, the local range variation (LRV) is computed. Very small regions including single points are classified as non-ground points basically due to the outlier behavior of these points. All other regions are evaluated by the LRV descriptor. Regions having height jumps above a certain threshold on their boundaries will be evaluated as non-ground regions. The LRV values of each boundary region are extracted. If the majority, i.e., here 90 percent, of the LRV values are above a threshold, here $0.5m$, the region is classified as non-ground. For the remaining regions the discontinuity (or slope) along the boundary is supposed to be not significant, and thus they are

considered to be terrain regions.
The steps from creating the *marker* image to the classification of the non-ground regions are repeated for all *marker* images. The classification takes the result of previous iteration into account by merging it with the classified non-ground regions of the current iteration. One

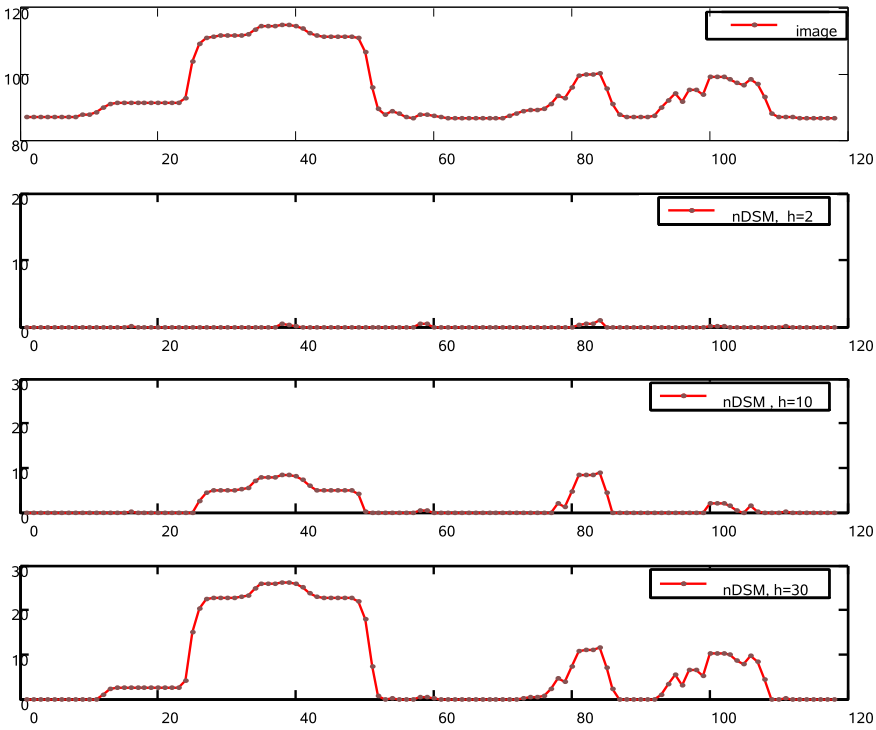
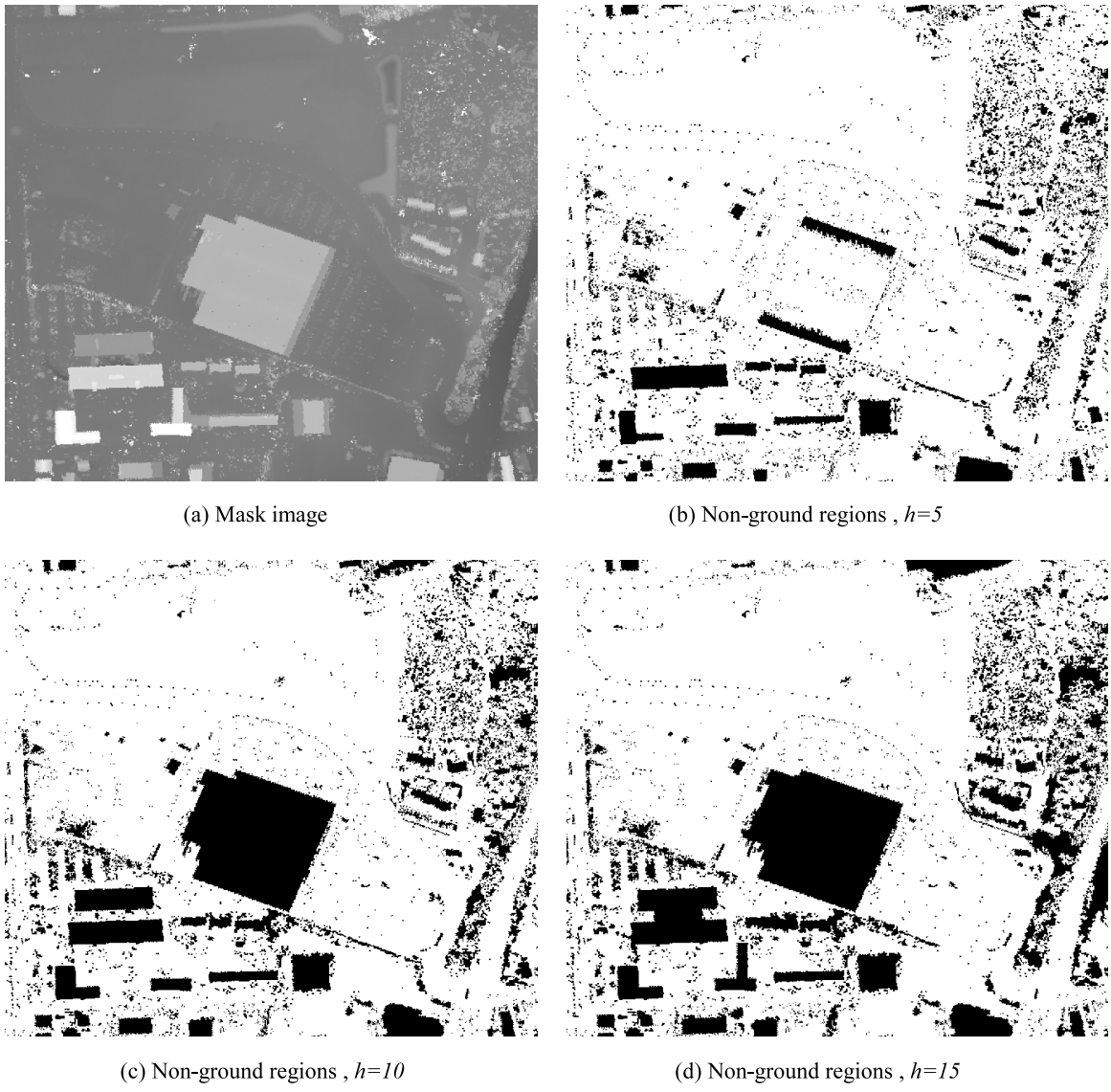


Figure 4.11: Different $nDSMs$ based on grayscale reconstruction. From top to bottom: *mask* image, $nDSM_0$ with $marker = Mr - 2$, $nDSM_0$ with $marker = Mr - 10$, and $nDSM_0$ with $marker = Mr - 30$

aspect of the proposed algorithm namely using a sequence of marker images rather than just one marker image has to be discussed in more detail: The impact of employing different height offsets when creating marker images is shown in Figure 4.11. With small height offsets only non-ground regions with a low height are addressed. Figure 4.12 shows that some of these regions are representing only a part, mostly the top, of larger regions. By increasing the height offset also high non-ground regions are considered. Not necessarily are the high non-ground regions also larger. But in contrast to the *TopHat* filtering approaches (cf. Section 2.3.1), which filter the data with different kernel sizes to cope with objects of different size, the extent of an object does not matter for our approach. Using a sequence of marker images provides for all structures the appropriate height offset level at which each of the non-ground regions has the maximum height discontinuity along its boundary. In this way non-ground regions with a variety of different sizes and heights can be extracted.

4.3 Generation of Normalized DSM and DTM

In the previous section a method based on geodesic image reconstruction was employed to classify the laser data into ground and non-ground regions. The next step is to generate a normalized DSM or $nDSM$ from the original laser data. For this, the ground data is re-interpolated and the gaps produced by filtering are filled. Contrary to the interpolation of the



(a) Mask image

(b) Non-ground regions , *h=5*

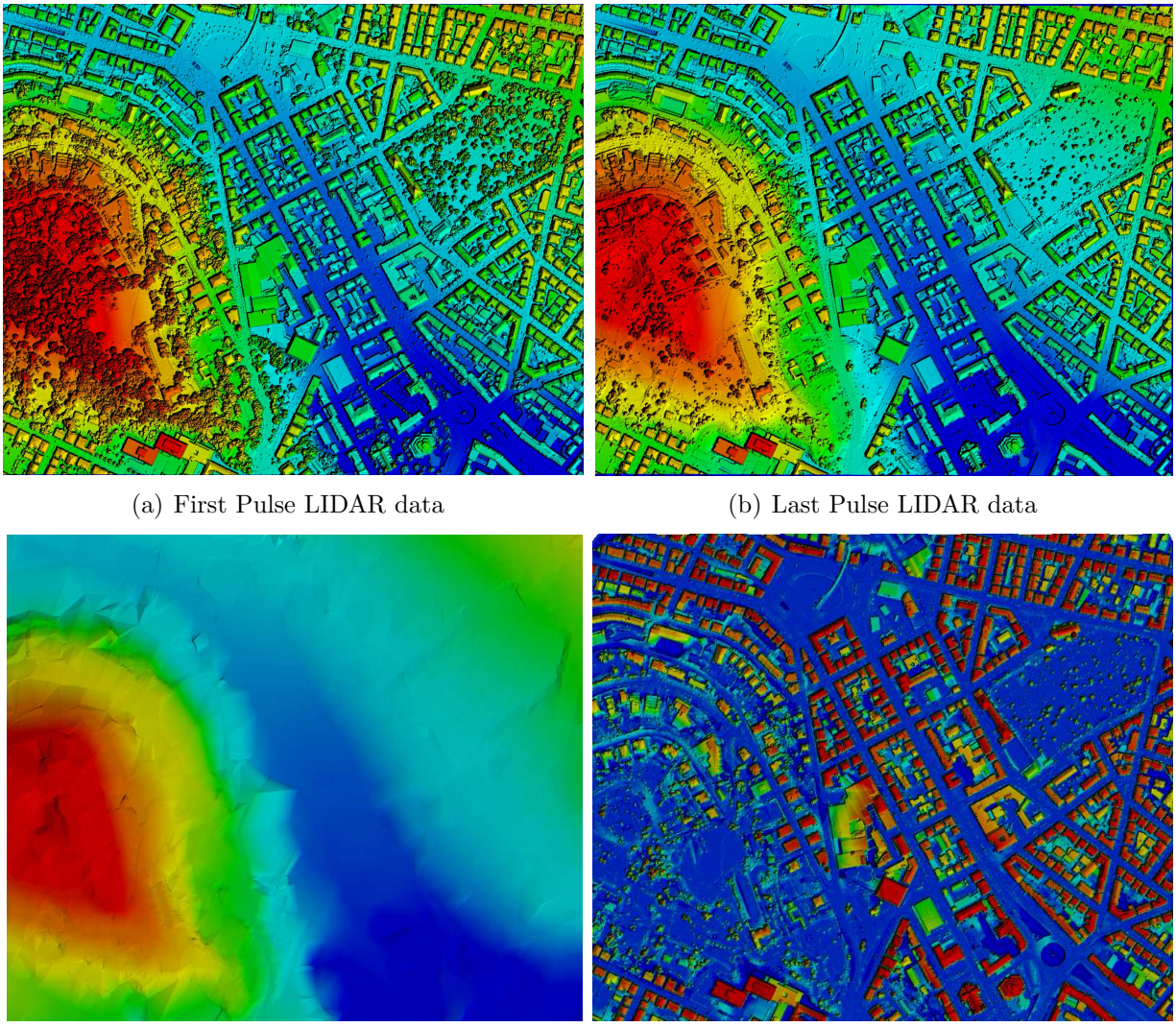
(c) Non-ground regions , *h=10*

(d) Non-ground regions , *h=15*

Figure 4.12: Non-ground regions detected by applying different height offsets (Black color represents the non-ground regions). Integration of the non-ground regions detected based on each height offset provides the final result.

raw laser data, here an Inverse Distance Weighting (IDW) algorithm is used. The result is a smoothed surface which is basically a Digital Terrain Model (DTM). Afterwards, the DTM is subtracted from the original LIDAR data to generate the $nDSM$.

Figures 4.13(a) and 4.13(b) represent first pulse and last pulse range images, respectively, in a hilly residential area. Small buildings as well as spacious building blocks exist in the area. A big part of the area is covered by dense vegetation located around the top of a hill. The difference between the minimum and the maximum height values is about $204m$ in a region of about $1km \times 1km$.
The feature descriptor which highlights the local height changes, i.e., LRV, is represented in Figure 7.6(b). After detecting all non-ground regions from the LIDAR image, the pixels belonging to them are eliminated from the image. The gaps are later filled using interpolation to generate the filled DTM (cf. Figure 4.13(c)). Figure 4.13(d) illustrates the nDSM generated



(a) First Pulse LIDAR data

(b) Last Pulse LIDAR data

(c) DTM generated from LP data

(d) Normalized DSM for LP data produced in final step

Figure 4.13: Airborne LIDAR data and result (area= $1\,km^2$, min height=$204.5\,m$, max height=$407.5\,m$)

by subtracting the original image (cf. Figure 4.13(b)) from the DTM (cf. Figure 4.13(c)). The nDSM is produced from the last pulse image. As shown in this Figure, all the 3D objects are located on a background with *zero* height level.

Figure 4.14: Local Range Variation (LRV) image which highlights height jumps

Chapter 5

Detection and Approximation of Building Outlines

In this chapter the proposed approach for the detection and approximation of the building outlines is discussed. it assumes that an $nDSM$ is given by means of the geodesic image reconstruction algorithm explained in the previous chapter. The proposed process for the classification of LIDAR images into buildings and trees is illustrated in Figure 5.1. The algorithm contains three steps: Segmentation of non-ground regions, extraction of feature descriptors, and rule-based classification. In the following sections, all three steps are explained in detail.

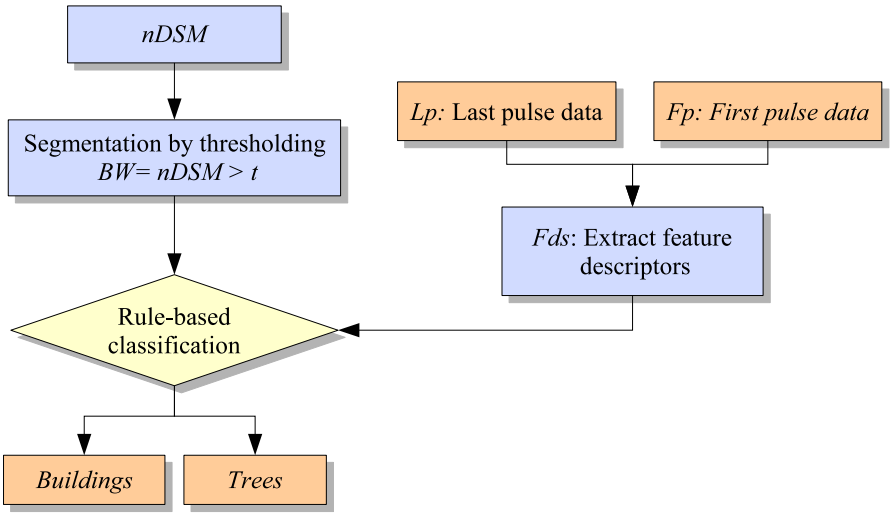


Figure 5.1: Classification of LIDAR range data into buildings and trees

After the classification the proposed method for approximation of the building outlines is described. The goal is to reduce the number of points of the building outline and to produce a rectangular polygon if possible. For this two different approaches are proposed: One fits a rectangular (*rectilinear*) polygon to the building outline by hierarchical rectangle fitting. In the second approach the building outline is approximated by fitting straight lines and merging them to generate proper polygons.

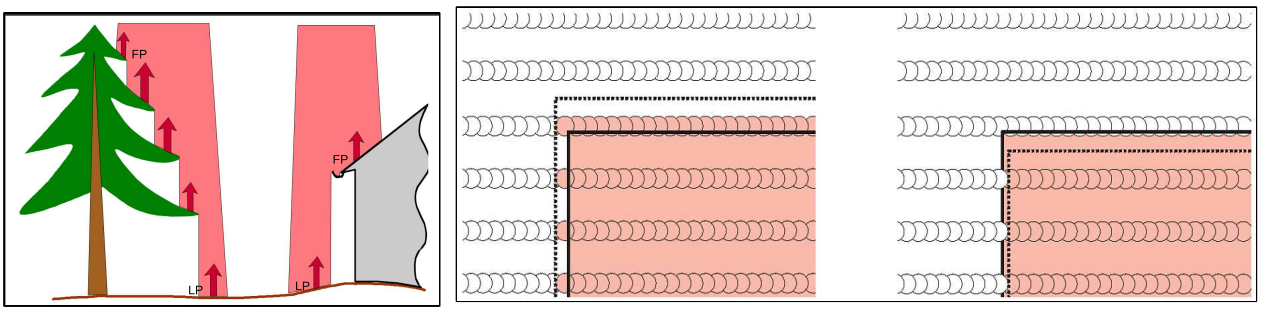
5.1 LIDAR Data Classification

The hierarchical process for the extraction and classification of 3D non-ground regions from LIDAR data starts with the normalized DSM ($nDSM$), the major outcome of applying geodesic image reconstruction to LIDAR range data (cf. Chapter 4). The $nDSM$ is thresholded to separate potential non-ground regions. Region segmentation, classification of the regions based on region properties, and fusion of the regions into one final result are further steps. The classification of the regions is carried out rule based and uses geometric region properties together with other feature descriptors. The properties comprise the size of the region, the vegetation index VI based on first and last-pulse range data, the local height variation defined by the difference between the maximum and the minimum height in a local window, and the variance of the surface normal. The proposed process is illustrated in Figure 5.1 .

5.1.1 Extraction of Feature Descriptors

In the following we discuss the computation of feature descriptors including filtering in a preprocessing step which comprises the following feature descriptors:

1. Calculation of the VI image which is employed for the separation of vegetation points from non-vegetation points. A main property of the ALS systems is that the laser pulses partially penetrate the vegetation in gaps between leaves and collect data from points underneath the vegetation. While in first-pulse (FP) mode data on the vegetation's surface is available, this is not the case in last-pulse (LP) mode(cf. Section 2.1). Another effect of multiple return measurement occurs when measuring an object with a large height discontinuity, such as building. In such objects if the laser footprint is locates on discontinuity, e.g., a building edge, the FP measures a point on a higher elevation level than the LP (cf. Figure 5.2). Therefore, the same building may have larger dimension in FP data compared to a LP data (cf. Figure 5.2(b)) or a point in FP data located on the building edge may have a lower height value in the LP data (Steinle and Vögtle, 2000; Clode et al., 2004). Figure 5.2 sketches the effect of multiple return laser measurements on the building boundary. The actual building outline is drawn by a continuous black line in FP (left) and LP (right) data. The small circles represent the laser footprints and dashed lines show the acquired contour lines in FP and LP data. The images prove that the building dimension in FP data is larger than the actual one while it is smaller in LP.



(a) Multiple return

(b) Changes in building dimension in first- and last-pulse images (Steinle and Vögtle, 2000)

Figure 5.2: Effects of multiple return measurements on 3D objects

Figure 5.3(c) displays an image VI_0 provided by subtraction of the LP (cf. Figure 5.3(b)) from the FP (cf. Figure 5.3(a)) image. An inspection of the bright pixels shows that the building edges are often highlighted as well as the vegetation. The behavior is similar for other objects such as powerlines and cranes. Figure 5.3(d) shows a vegetation index VI which is provided by subtracting the grayscale dilation of the LP from the FP image. A grayscale dilation using a small SE, e.g., a disk with a radius of 3, increases the dimension of the objects in the image. Therefore, the produced VI image does not contain the edges any more (cf. Figure 5.3(d)). The dilation on the LP image produces some negative pixel values in the VI image. They are converted to zero because they don't play any role for vegetation extraction. Figures 5.3(e) and 5.3(f) visualize the vegetation indexes provided in Figures 5.3(c) and 5.3(d) as RGB images. In these images the FP, LP, and vegetation index images are allocated to the blue, green, and red channel. Comparison of these images shows, the VI image produced by the second method can be a better representative for vegetation objects than the first method.

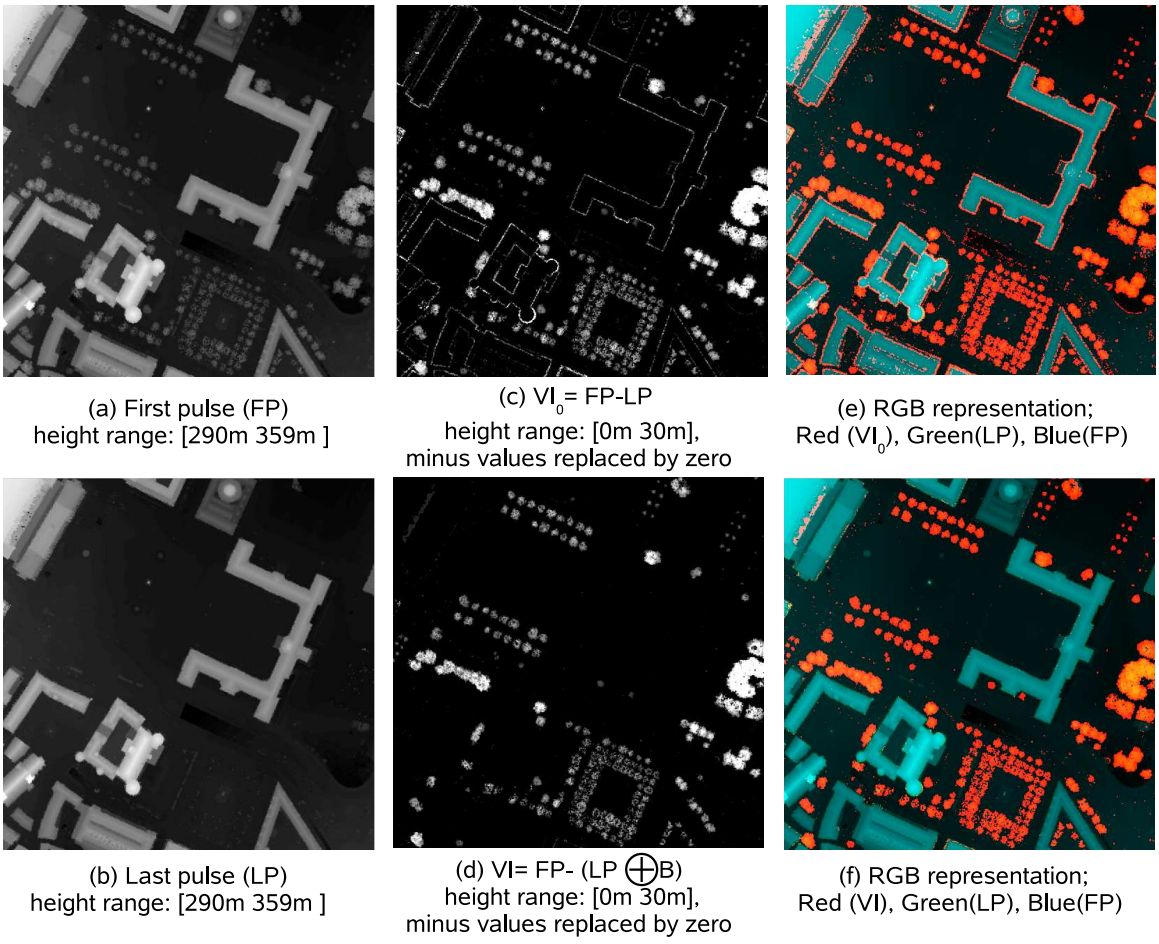


Figure 5.3: Generation of Vegetation Index (VI) feature descriptor from first-pulse (FP) and last-pulse (LP) images

2. Computation of the Local Range Variation LRV image as second textural image. It serves for separating high elevated objects, e.g., buildings and tress, from the other objects such as small bushes and ground points from last pulse image. It is assumed that most of vegetation regions are already recognized by employing previous feature descriptor, e.g., VI. Therefore, here the focus is only on last pulse image and detecting high elevated

buildings and trees by employing a bigger threshold for binarization as shown in Figure 5.1.

The LRV feature consists of the difference between the maximum and the minimum height value in local 3×3 windows in the LP image. LRV is used to evaluate the boundaries of the segmented regions (cf. Figure 7.6(b)).

3. A third feature descriptor is the variance of the surface normals. This hight texture is a homogeneity criterion and used as one of the common methods for separating trees and buildings from DSM (Weidner and Förstner, 1995; Brunn and Weidner, 1997).

   Here, it is employed for refinement with the goal to detect remaining vegetation areas from last pulse image, i.e., those not indicated properly by the VI, and separate them from building areas. Geometric region properties like size or shape are not sufficient for separating these two types of regions, because, e.g., large areas with rectangular shape may indicate vegetation as well as buildings.

   As alternative to the variance of the surface normals the other features such as Laplace filter (Maas, 1999a; Vögtle and Steinle, 2003) or local curvature (Vögtle and Steinle, 2003) can be used for this purpose.

5.1.2 Segmentation of Non-ground Regions

The $nDSM$ filtered laser data as well as the VI image are binarized by thresholding. For the $nDSM$ a threshold of, for example, $2m$ is used to separate objects from the ground. For the VI a threshold of $30cm$ is employed to distinguish vegetation from the ground. The thresholds directly reflect the assumptions about what is considered as a non-ground point. Connected components are determined and labeled to obtain regions. For each individual region properties like the size of the region are calculated.

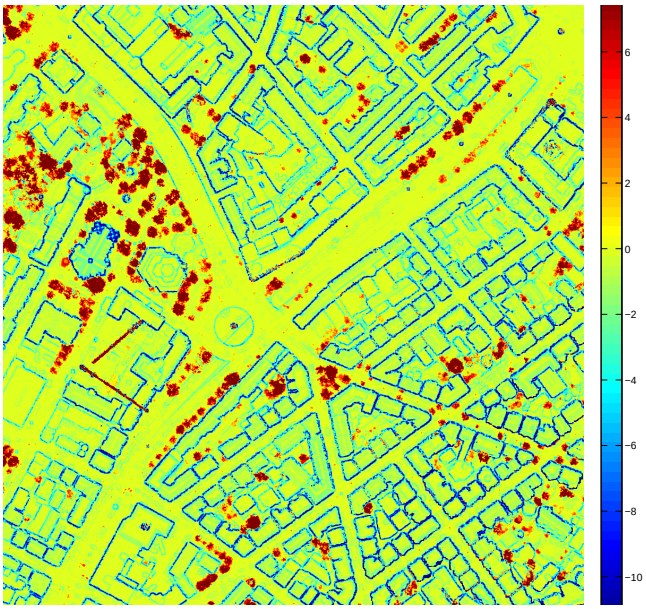
5.1.3 Rule-based Classification

A number of decisions is made according to the following rules:

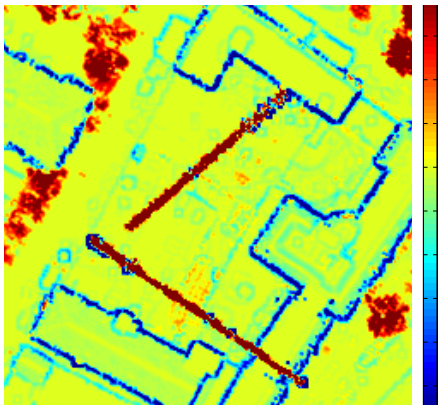
- **Decision 1:** Regions are considered to be *mostly* vegetation regions if the VI values indicate an average height difference between FP and LP range of more than $30cm$ (cf. Section 5.1.2). The threshold takes into account measurement noise and meaningless small height differences at the object. In reality there is a number of objects which show almost the same behavior as vegetation in LIDAR data. This means, they are visible in the FP data and not visible in the LP. Power lines and cranes are such objects.

  Figure 5.4(b) shows two cranes as well as vegetation regions highlighted by subtracting the FP range from the dilated LP data. The cranes are shown in dark red color and located almost center-left of the image. They are almost $30m$ and $45m$ higher than the neighboring ground pixels. As illustrated, the crane pixels are highlighted very similar to the vegetation pixels (cf. Figure 5.4(a)). The following strategy is employed to discriminate the vegetation pixels from the crane pixels:

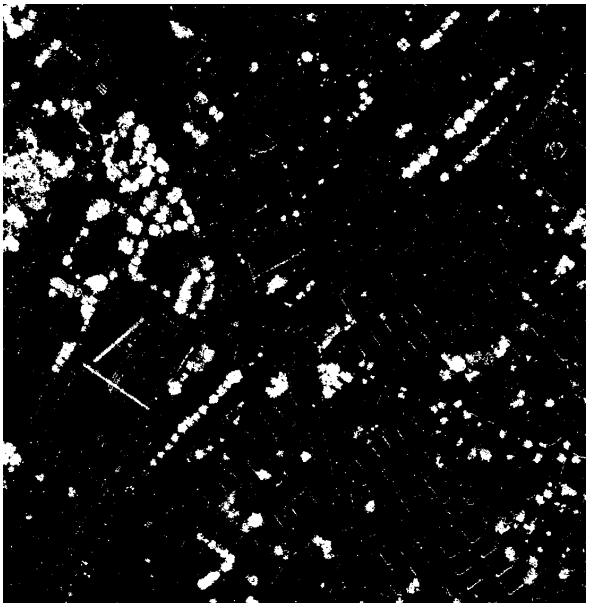
  Contrary to the vegetation, the cranes (here, the higher horizontal part of what is visible from top) and power lines contain a very large height variation in range image. They are linear, narrow, and horizontal, i.e., the pixels have almost the same height value.



(a) Highlighted regions by subtracting the dilation of the LP data from the FP data. The minus values in the image are produced by employing image dilation instead of original LP data



(b) Two cranes in detail view

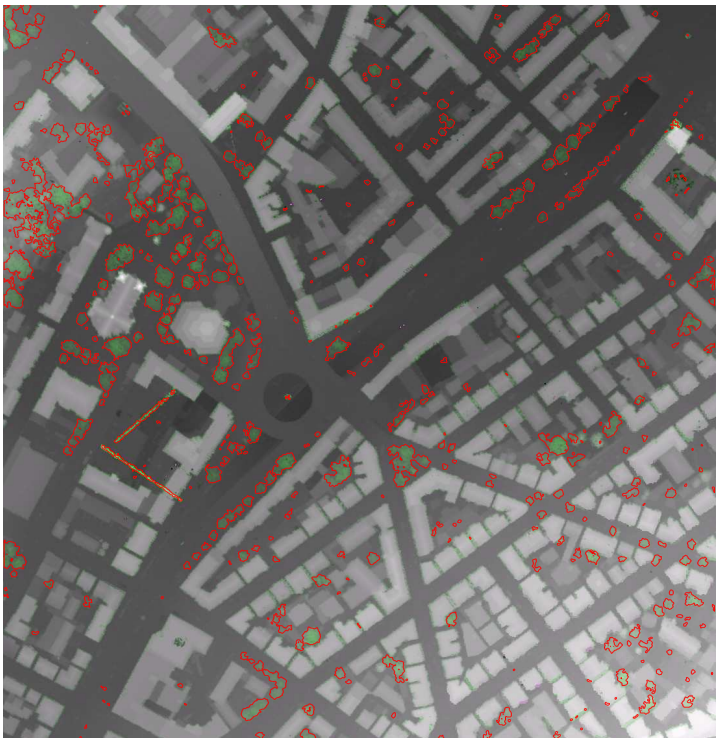


(c) Binarized regions by employing a threshold value of $t > 30cm$

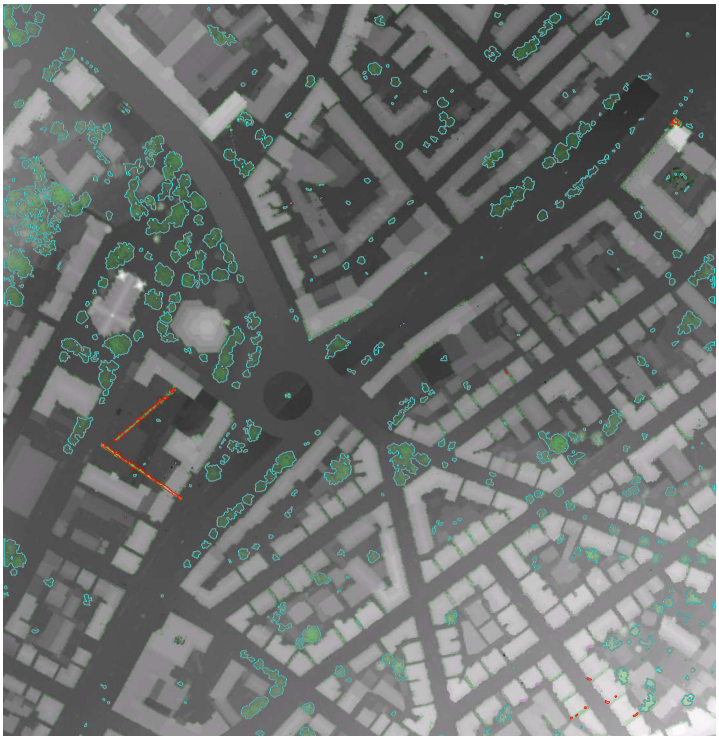
Figure 5.4: Detected regions by subtracting the dilated last-pulse from the first-pulse data

Therefore, the variation of the heights are greater compared to the vegetation pixels. To separate those regions from the vegetation regions, the variance of the height values can be computed and evaluated. For that, the raw Z values, i.e., before interpolation, corresponding to each region are extracted and the variances evaluated. Figures 5.5(a) and 5.5(b) represent the result of vegetation extraction before and after evaluation of the height variances.

To detect the building regions, all the regions considered to be vegetation, cranes or power lines by this decision are eliminated from the set of regions.



(a) Detected regions before considering the height variations *(red polygons)* superimposed on the original FP data



(b) Refined vegetation regions *(blue polygons)* after considering the height variations to separate objects such as cranes and power lines from the vegetation regions

Figure 5.5: Final vegetation regions extracted by taking into account the vegetation index VI feature descriptor. The cranes are separated from vegetation regions by evaluation of the variances of the pixel values (heights) inside each region.

- **Decision 2:** The remaining regions are further investigated based on the LRV image. Only the boundary pixels of each region are taken into account. Regions that have an average height variation along the boundary of more than, e.g., $2m$ (cf. Section 5.1.2) are seen as as non-ground regions. Otherwise, those regions are considered to represent the ground.

- **Decision 3:** The remaining regions mostly represent buildings, but still some vegetation areas are present. The discrimination between vegetation and building regions is based on the variance of the surface normals calculated for all pixels within the boundary of each region. Regions with a low value for the average of the variance of the surface normal are classified as building regions, the remaining ones as vegetation regions (cf. Figure 5.6).

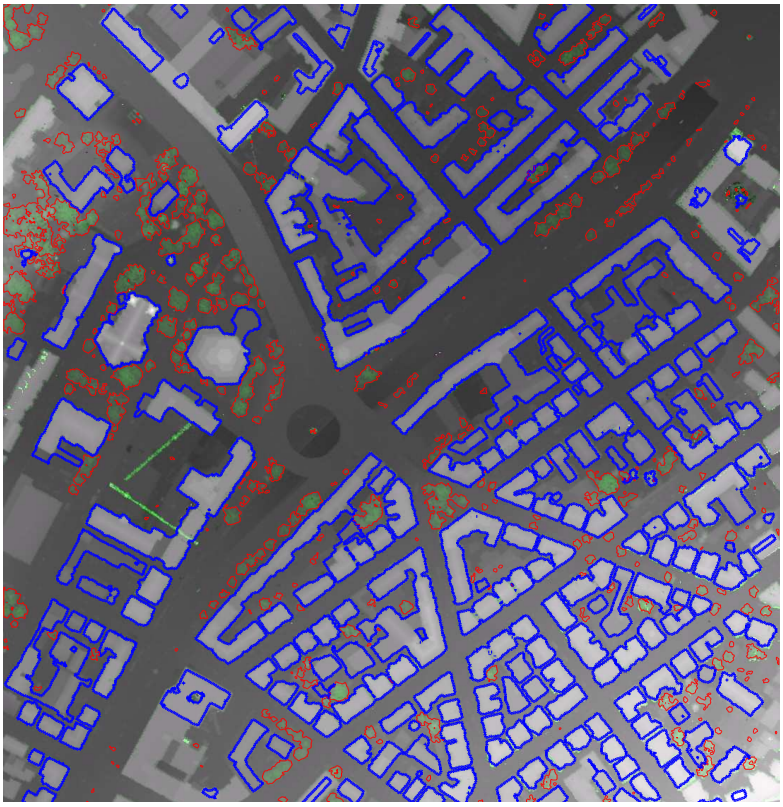
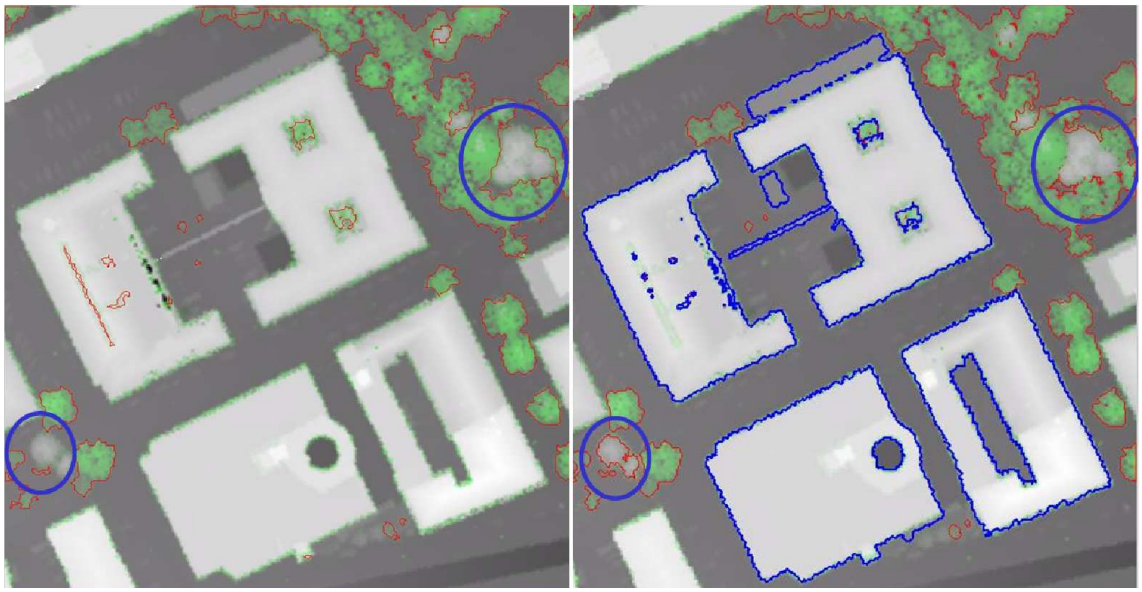


Figure 5.6: Building *(blue polygons)* and vegetation *(red polygons)* detection output

Figures 5.7 and 5.6 show building and vegetation classification results by means of the rule-based approach. For a better visualization, a combination of LP and FP is used to generate an RGB image. The green pixels highlight $FP - LP$ and are mostly tree pixels. The green pixels along the building boundaries proves that the building size in FP is often larger than in LP (cf. Section 5.1.1). A few green pixels are also located at glass roofs where $FP - LP$ is not *zero*. Some trees are not highlighted with green pixels, e.g., trees with gray colors inside blue circles. Those trees contain dense leaves which cause the same reflection in FP and LP images, i.e., $FP - LP = 0$.

Figures 5.7(b) and 5.6 represent the final classification result superimposed on the original FP image. Building and vegetation polygons are displayed as blue and red polygons respectively.



(a) RGB image produced from LP and FP images; regions detected from the segmentation of vegetation index (VI) image overlaid as red polygons.

(b) Classification result for buildings (blue polygons) and vegetation (red polygons)

Figure 5.7: Building and vegetation detection; a) highlights the vegetation regions detected directly from the vegetation index *(VI)* image. Blue circles show the areas where the vegetation is not detected; b) shows the final classification for buildings and vegetation. In this Figure the vegetation regions detected rule-based are integrated with the vegetation detected from *VI*.

In this Figure the vegetation regions detected from rule-based method are integrated with the regions extracted from the VI image.

5.2 Building Outline Approximation

After extracting the building regions, the outline of the buildings needs to be extracted and simplified before 3D modeling. The building polygons often contain many unnecessary vertices which can be reduced without a big effect on the shape of the building.
In general, polygon approximation, also termed polygon simplification, is a process to approximate an input polygon by an output polygon with fewer vertices, while maintaining as much of the initial morphology as possible.
Buildings are often constructed using relatively simple ground plans consisting of straight lines and right angles. The inclusion of the original building boundaries derived above into a GIS would add unnecessary complexity. Furthermore, simple ground plans form a suitable basis for the 3D modeling of buildings. All this leads to the following goal: Find a polygon as simple as possible fitting to the original boundary points as good as possible. In spite of the fact that some building boundaries are not straight, it was decided to approximate all sides of the boundary polygons with straight line segments. Thus "simple as possible" means "having as few corner points as possible". The requirement "fitting to the original boundary points as good as possible", however, can be interpreted differently and is stated more precisely below.
As neither the approximate coordinates of the corner points nor their number is known, the problem stated above is not merely a problem of parameter estimation, but additionally a problem of finding, refining and testing of model hypotheses. In the following requirements for fitting a polygon to a building outline are given (Arefi et al., 2007a):

1. Significant parts of the building should be preserved.
2. "Very small" structures or "very short" edges should be eliminated.
3. Inherent symmetries of the building should be preserved.
4. The number of the resulting linear segments should be small. In particular, there should be no "short" neighboring line segments with an angle of close to 180°.
5. The distances between the given points and the polygon should be small, e.g., in the sense of least squares.
6. The normal vector of the resulting polygon segment should coincide as good as possible with the normal of the given boundary.
7. The size of the areas enclosed by the original points and the polygon should be as similar as possible.
8. The developed methods should be robust against outliers.
9. It should be possible to control the refinement of the model, or the degree of generalization, respectively, by means of easily interpretable parameters.

Whereas the requirements 5-7 can be immediately expressed in terms of mathematical formulae, requirements 1-4 are less well defined. While characteristic features of buildings such as the size of the area can easily be formalized and thus be preserved, this is not the case for features like symmetry. Furthermore, the criteria for a good approximation depend very much on the individual application. For instance, the following additional requirements may exist:

- Angles close to 90° (or 270°) should be set to 90° (or 270°).

- All angles of the polygon should be set to 90° (or 270°) in case of rectilinear polygons.

In principle, two classes of solutions are possible: Either a simple polygon, for instance a rectangle, is fitted to the ground plan and then iteratively refined until the requirements are fulfilled as far as possible, or linear segments are fitted to the boundary points and combined, e.g., intersected, supplemented by interpolating segments, etc., to form a closed polygon. The former solution suits for buildings with rectilinear outlines, i.e., right angles in all corners, while the latter is appropriate for all buildings with straight line edges.

Representatives of the first class which hierarchically fit polygons to the points are:

1. Iterative Construction of Minimum Bounding Rectangles (MBR), i.e., hierarchical fitting of rectangular polygons (Wikipedia, 2007): This method which is explained in detail in section 5.2.1 is a top-down, model-based approach that hierarchically optimizes the initial rectilinear model by fitting MBR to all details of the data set.

2. The Ramer or Douglas-Peucker algorithm (Ramer, 1972; Douglas and Peucker, 1973): It starts with the first and last point A and B. A straight line is constructed through these points. The original point Q having the largest distance from the straight line is determined and the straight line is replaced by the two line segments AQ and QB if the distance is beyond a given threshold. This is repeated recursively until the distance between the polygon and the given points is below the predefined threshold. For a closed polygon one selects one point randomly and chooses as second point the point with the largest distance from it. Contrary to the technique mentioned in point 1 this method doesn't produce rectilinear polygons.

In contrast to the methods of the first class which are top-down, in the second class where the straight line segments are determined first, the techniques are bottom-up. In top-down methods the initial model, i.e., polygon, is shaped first and then optimized iteratively, while in bottom-up techniques the details, i.e, line segments, are determined first and then merged afterwards to form the final model.
The following well-known methods can be applied in order to fit straight lines to point clouds being members of the second class of solutions:

1. Hough Transform (cf. Section 2.4.1)

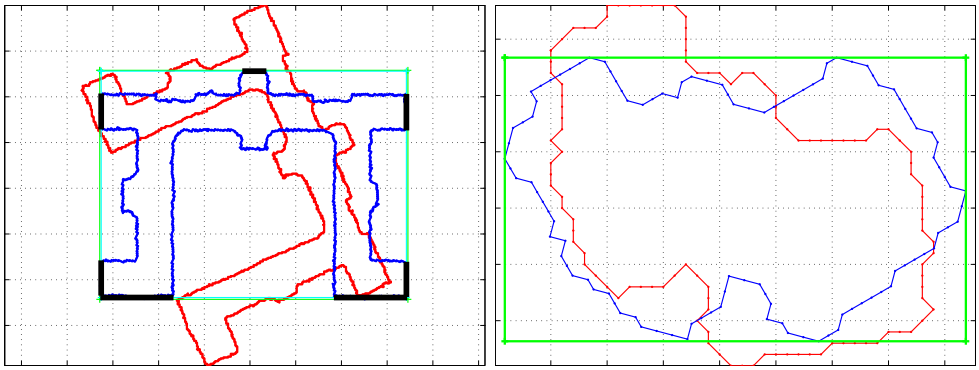
2. RANSAC (cf. Section 2.4.2)

Both above cases are supplemented by algorithms integrating the line segments into a closed polygon.

5.2.1 Boundary Approximation by Hierarchical Fitting of Rectangular Polygons

- **Determination of Main Orientation:** The method for boundary approximation described below assumes that the coordinate system is oriented parallel to the main orientation of the polygon. I.e., most polygon edges are parallel to either the X- or the Y-axis of the coordinate system. To determine the main orientation, two methods are proposed and evaluated: one is based on *the minimum bounding rectangle*, the other on the *Hough transform*.

- **Minimum Bounding Rectangle (MBR):** The MBR is computed by the classical Rotating Calipers algorithm (Shamos, 1978). The original polygon is sequentially rotated in small steps around its center of gravity, e.g., in steps of 0.5 degree. In each step the bounding box is determined and its area is calculated. The rotation angle which results in the minimum bounding box defines the main orientation of the polygon.



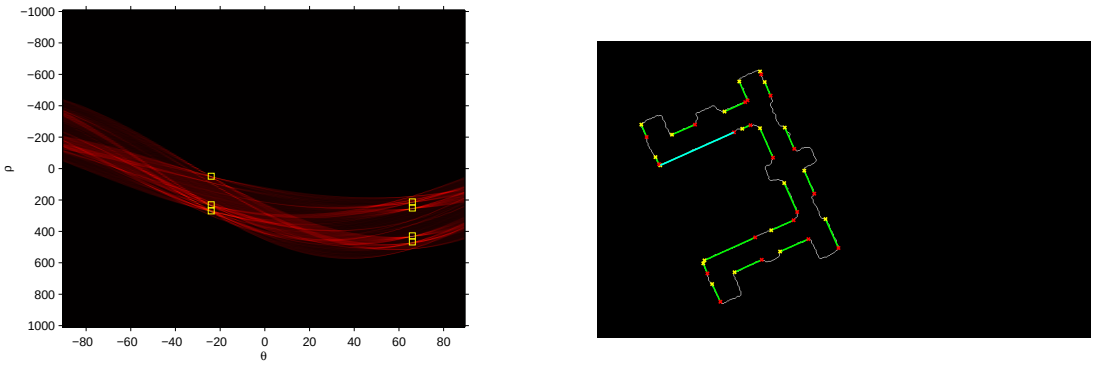
(a) Example1 (b) Example2

Figure 5.8: Main orientation using MBR

In Figure 5.8 the red polygon represents the original building boundary, the blue polygon the building boundary rotated parallel to the main orientation, and the green rectangle the MBR of the rotated polygon. Figure 5.8(a) shows that the method properly determined the main orientation of the polygon, i.e., the edges of the rotated polygon (blue lines) are parallel to the X- or Y-Axis of the coordinate system. Figure 5.8(b) illustrates the obtained MBR for another object. It makes clear that the MBR rectangle is not always parallel to the "main edges" of the building.

- **Hough Transform:** As the main orientation found by the MBR is not always parallel or perpendicular to the direction of the edges, another method based on the Hough transform is devised. In this method the main orientation is determined based on the average of the orientations of the Hough lines (cf. Section 5.8(a)) weighted according to their lengths. The Hough transform results in a histogram for the described edges' angles (X- axis) and their signed discretized distance from the origin (Y- axis). The angle for which most edges provides evidence in the form of the sum of contributions over the respective column as well as of the 90° shifted column is assumed to correspond to the main orientation of polygon.
  In practice, the Hough Transform is computed for the edge points and the main Hough peaks, e.g., the best 10 peaks, are extracted. The corresponding angles for selected peaks are checked if they are complementary with others. If the difference between two peaks is a multiple of 90^o with an accepted small deviation, only one of them is kept. The angle which has the maximum support is selected as main orientation. The final list of remaining angles shows whether there is only one independent angle, i.e., not complementary with others, as main orientation, or more.

  In Figure 5.9(a) the positions of the main peaks are marked by yellow squares in the Hough transform of the boundary image. Lines corresponding to the detected peaks are highlighted in Figure 5.9(b). Figure 5.10 shows that the orientation by means of the Hough Transform is also correct for the example 2 of Figure 5.8(b), where the method failed based on the MBR.



(a) Hough peaks

(b) Lines corresponding to peaks of the Hough Transform

Figure 5.9: Main orientation by Hough Transform

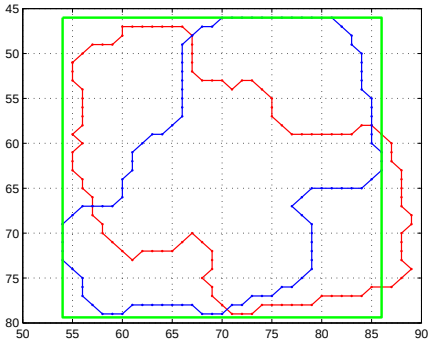


Figure 5.10: Main orientation by Hough Transform for example presented in Figure 5.8(b)

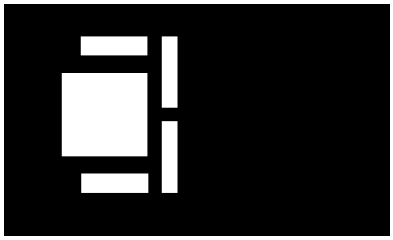
- **Hierarchical Fitting of Rectangular Polygon:** After determining the main orientation, the building polygon is rotated to the main orientation, as shown in Figure 5.11(a). In the next step, the MBR image (cf. Figure 5.11(b)) is subtracted from the rotated building region (cf. Figure 5.11(c)). The subtraction produces new regions, cf. Figure 5.11(c). For any of those regions a MBR will be calculated (cf. Figure 5.11(d)). They are again subtracted from their corresponding regions produced in the previous step (cf. Figure 5.11(e)). As illustrated in Figure 5.11(e), some small regions are created. The process is repeated by again computing MBR regions and subtracting them from their corresponding regions. This hierarchical procedure is continued until either no new regions are created any more or the size of the produced regions is less than a predefined threshold. After convergence, the polygon is rotated back to the original orientation, cf. Figure 5.12 with the red lines highlighting the rectangular polygons.

The above procedure employed for the outline of each building in the image. For buildings containing courtyards, the points of the outer and the inner boundaries are individually processed and corresponding rectilinear polygons are extracted. The final approximated polygon is formed by subtracting the courtyard, or inner boundary points, from the outer boundary points. An example for the approximation of buildings with courtyards is presented in Section 7.3.3.

(a) Rotated building

(b) MBR image

(c) Rotated region - MBR region



(d) MBR on small regions

(e) New regions produced by subtraction of Figures (5.11(c) and 5.11(d))

Figure 5.11: Segmentation based on MBR

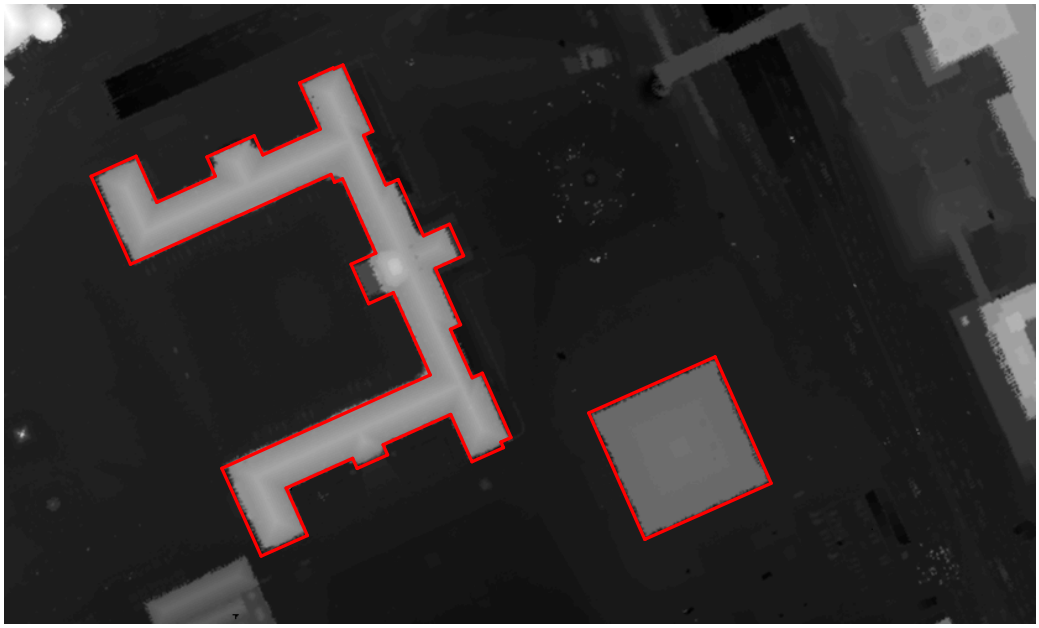


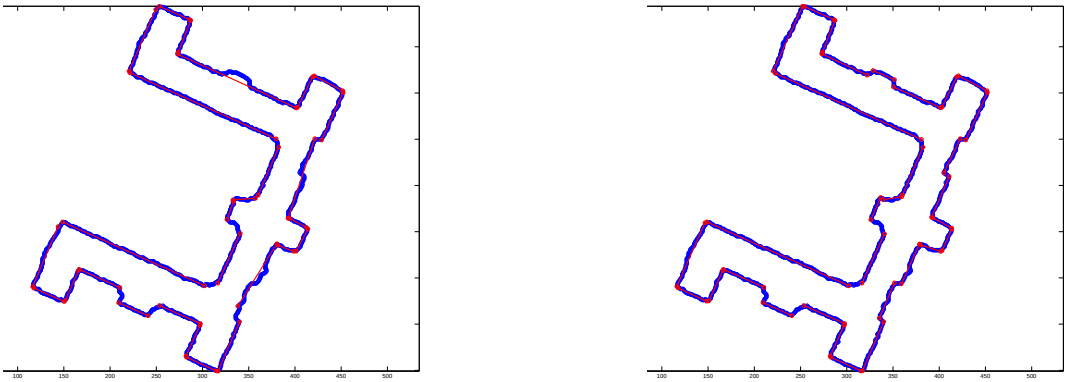
Figure 5.12: Final rectangular polygons superimposed on original LIDAR image

5.2.2 Boundary Approximation Using RANSAC

RANSAC (cf. Section 2.4.2) was originally devised to robustly fit *one single* model to noisy data. It turns out, however, that it can also be successfully used in a procedure to fit a *beforehand unknown number* of models, as in the case of the ground plan boundaries the number of line segments is initially unknown. The RANSAC-based approximation method proceeds as follows: Straight line segments are repeatedly detected after deleting already fitted given points from the input data until either a) the lines found so far are considered sufficient to construct the ground plan completely or b) the number of points fitted to the best line segment with respect to the current iteration step falls below a chosen threshold t (Arefi et al., 2007a). This threshold already appears in the "original" RANSAC: "The RANSAC paradigm contains three unspecified parameters: (1) the error tolerance used to determine whether or not a point is compatible with a model, (2) the number of subsets to try, and (3) the threshold t which is the number of compatible points used to imply that the correct model has been found" *(ibidem)*. To adapt RANSAC for the particular problem, a modification is suggested which is reducing the computational effort: If pairs of points are selected randomly as by the original RANSAC, for a lot of pairs the two points would belong to different line segments. By making use of the fact that the points are given in the order they are connected, only one point is selected randomly and the second point is chosen from a neighborhood of the first defined by a specified index difference Δi to the first point. Δi is a further parameter of the method. If it is chosen too small, the accuracy of the line segment may be poor. If it is chosen too big, many pairs may result with the two points belonging to different line segments. Our experiments have shown that small values of the index difference are acceptable if the parameters of the line segment are determined iteratively. All points compatible with the line, i.e., the inliers, are used to compute improved line parameters, which again lead to a new improved set of inliers. This is repeated until no considerable improvement is obtained any more.

As the goal are straight line segments rather than lines, also the ends of the segments have to determined. For this purpose the outermost footpoints of the boundary points on the lines are computed. Sometimes, straight lines in the ground plan are interrupted, for instance by protruding parts of the building. Thus a further parameter s is introduced controlling how big a gap of non-fitting points is allowed to be inside a line segment. The line segment with most inliers is accepted for the respective iteration step. All boundary points between the first and the last compatible point are deleted from the input data.

The parameters s and t allow to control the degree of generalization of the model. However,



(a) $t = 15$, $s = 30$ (b) $t = 15$, $s = 7$

Figure 5.13: Line segments obtained by RANSAC – boundary pixels in blue, approximating line segments in red for different parameter settings (t: number of iteration; s: gap length)

large values for s may lead to undesirable results. Figure 5.13(a) shows on the right side two artificial skew segments, which are obviously supported by far-off points after gaps of still admissible length. Furthermore, there is a violation of the symmetry requirement: On top of the figure a small structure is eliminated, while on the bottom the corresponding structure is preserved. With the smaller value for s in Figure 5.13(b), a lower degree of generalization is achieved, but the symmetry of the building is reflected in the approximation.

When the iteration stops, because no more line segments with enough inliers are detected any more, the polygon is assembled from the line segments. The segments are sorted according to the ordering of their endpoints and the following rules are applied:

1. If neighboring segments are nearly parallel:
   (a) If the orthogonal distances of the corresponding straight line to the origin are nearly equal, the two segments are unified to one segment.
   (b) Otherwise, a new linear segment is inserted which is nearly orthogonal to both.
2. Otherwise:
   (a) If the distance between corresponding endpoints is small, the segments are intersected.
   (b) Otherwise, a new linear segment connecting the two endpoints is inserted.

These rules are obviously based on local information and thus result into local changes of the segments which have been detected by RANSAC. With a global adjustment exact 90°-angles

could be enforced and/or the area preserved. Yet, this would increase the computational effort. In summary, the following eight parameters have to be set for the described method (in the last column the values chosen in the present study are given – cf. Table 5.1):

	Par.	Meaning	Chosen Value
RANSAC	k	Number of random samples (nrp: number of remaining points)	$10 + nrp/100$
	d	Error Tolerance: Acceptable distance between boundary point and line segment	2 pixel
	t	Minimum number of fitting points for the acceptance of a segment	15
Gap Length	s	Admissible number of successive non-fitting points within a segment	7
Local Search Length	Δi	Index difference between the sample points	15
Assembling of segments to polygon	p	Parallelism criterion: angle between the normal vectors of two neighboring segments	18°
	u	Unification criterion: difference in orthogonal distances to the origin between lines corresponding to two neighboring segments	5 pixel
	e	Intersection criterion: Maximal distance between the endpoints of two neighboring segments	15 pixel

Table 5.1: Parameters controlling the boundary approximation using RANSAC

Figure 5.14 shows the polygon obtained using RANSAC. Only the cases "insertion of a line segment which is nearly orthogonal to both its neighbors" and "intersection of neighboring segments" occurred. There is an apparent asymmetry in the central part of the building: The skew segment was detected by the original RANSAC algorithm, the "orthogonal" one was inserted between its parallel neighbors. Such an asymmetry which already exists in the given data could only be avoided by either just connecting distant endpoints also for parallel segments, or adjusting short segments between parallel neighbors so that they become orthogonal.

5.2.3 Comparison of Methods for Building Outline Approximation

Two methods with very different characteristics have been presented for the approximation of building boundaries by straight lines. The MBR-based method is inherently top-down model-driven and hierarchical. It therefore can be adapted easily to different generalization levels. The approximating polygons form due to their orthogonality an excellent basis for the 3D modeling of building. A further important advantage is the good preservation of symmetries of buildings. The main drawback of the method is, that it is not suitable for ground plans with non-orthogonal edges. In this case the algorithm also converges, but only due to the

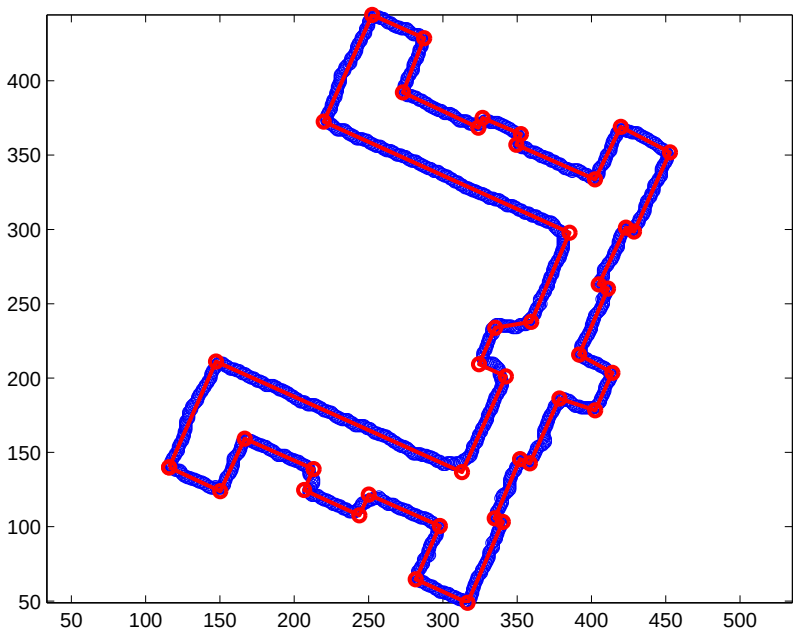


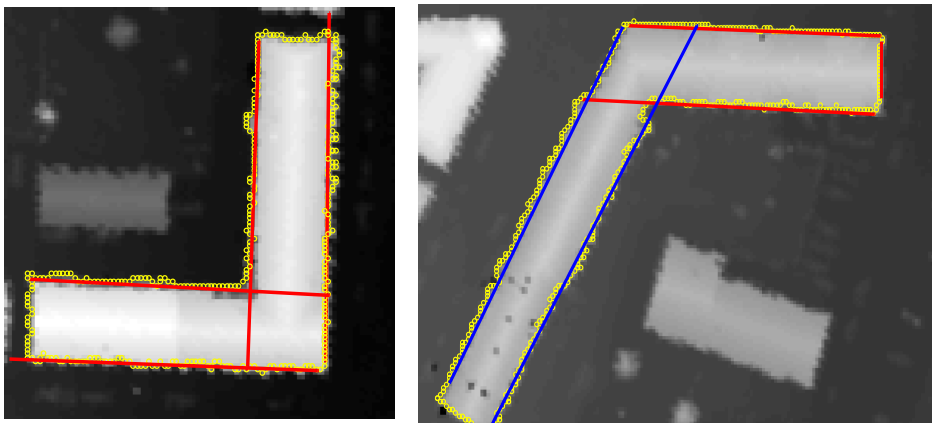
Figure 5.14: Approximation of polygon obtained using RANSAC

morphological Closing between the iteration steps. Saw-tooth patterns are generated whose amplitudes depend on the structuring element used for Closing.

In contrast, the RANSAC-based method forms approximating polygons by determining individual details, i.e., straight lines, and merging them. Since it does not rely on a main orientation, it is more flexible compared to the MBR-based method. On the other hand, the intermediate result, i.e., the set of line segments which are detected in the first step, depends strongly on the chosen parameters and is not reproducible since the points are selected randomly. For different initializations of the random number generator even the number of segments may vary. Additionally, an improper choice of the parameters may distort the symmetries of a ground plan. The rule-based algorithm which is used to combine the segments to a polygon also affects the result to a certain amount. However, a certain arbitrariness seems to be inherent in the problem, which becomes already evident from the unsharp requirements stated in Section 5.2. Therefore, both approaches have benefits and drawbacks. To make a general approach which is usable for all buildings with straight line boundaries, a combined method is proposed. It is based on number of main orientations defined by analyzing the Hough space of the building edges.

The buildings are classified into one main orientation (OMO) and many main orientation (MMO) buildings. Accordingly, OMO buildings are the buildings with orthogonal sides. In contrast, if in one building at least two neighboring sides are not perpendicular to each other, the building is many main orientation (MMO). If the analysis in Hough space indicates that there is more than one main orientation, i.e., MMO, the RANSAC-based technique is used (cf. Figure 5.15, right) otherwise the MBR-based method is implemented (cf. Figure 5.15, left).

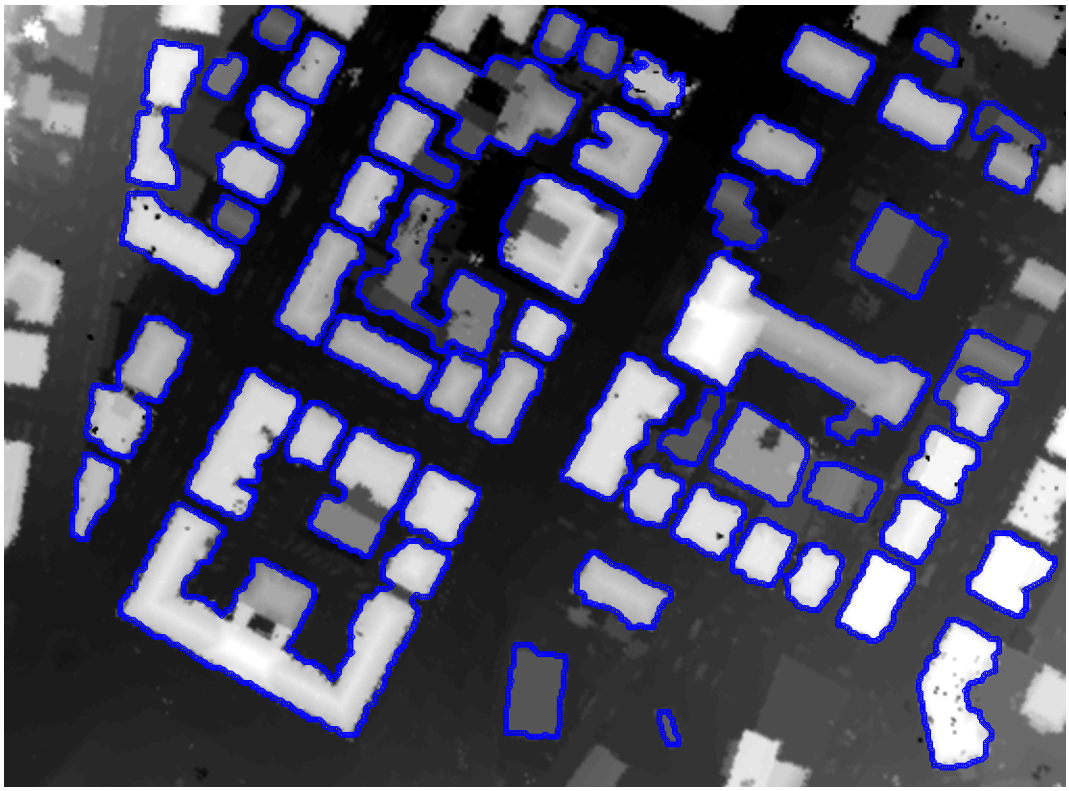
In Figure 5.15 the left building has a single main orientation, i.e., OMO, represented by the red lines, and the right building has two main orientations, represented by red and blue lines. Accordingly, outline polygons are extracted and approximated with the MBR- or the RANSAC-based method.



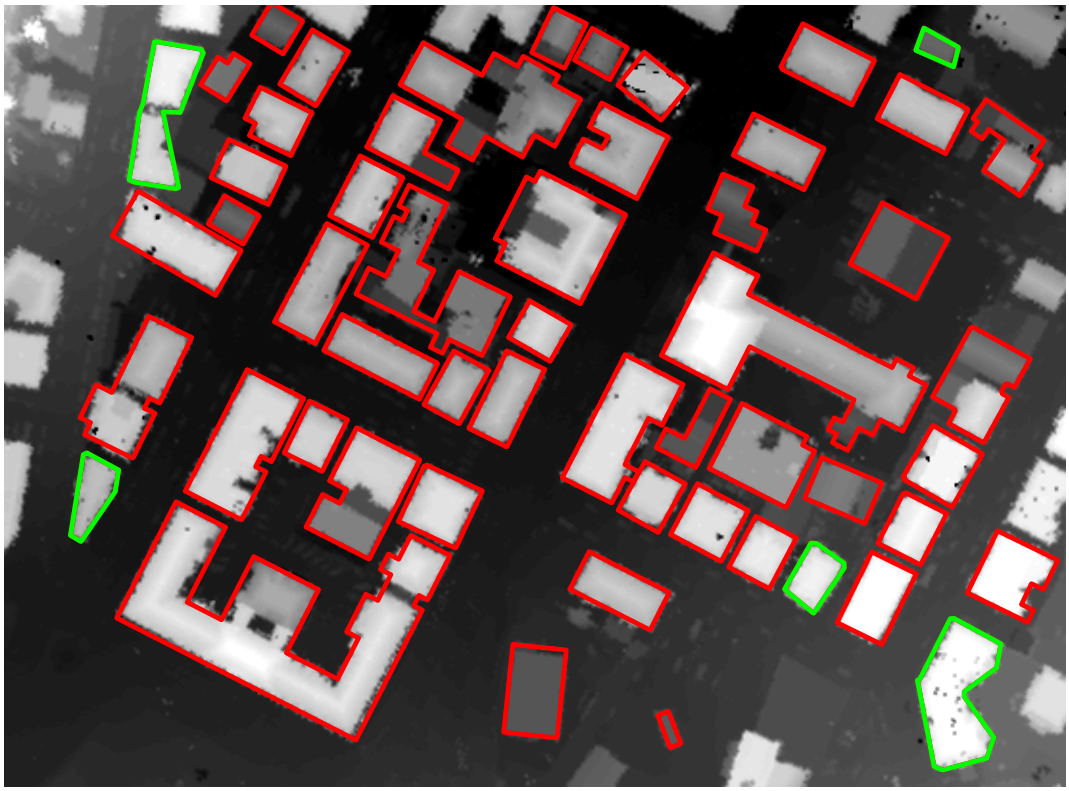
(a) One-orientation building (b) Two-orientation building

Figure 5.15: Building main orientations; The yellow points represents the building outline and the red and the blue lines are the lines found by Hough transform.

Figure 5.16 represents the result of the approximation procedure overlaid on the original LP image. The image contains many buildings with different sizes and orientations. The buildings displayed with red lines are classified as OMO buildings, i.e., Hough Transform found lines with orientations "almost" parallel or perpendicular to each other (cf. Figure 5.16(b)). These polygons are generated by means of the MBR-based technique as rectilinear polygons. When checking if the lines are perpendicular or parallel, a deviation of about $\pm 5°$ is taken into account, e.g, angles with an absolute difference of less than or equal to $5°$ are assumed to be parallel. Contrary to the OMO buildings, the MMO buildings are approximated using the RANSAC-based method (cf. Figure 5.16(b), green polygons).



(a) Building outlines superimposed on LP image



(b) Approximated polygons superimposed on original LP image; one main orientation buildings are approximated by rectilinear polygons (red) and two or more main orientation buildings are approximated using the RANSAC-based approach (green)

Figure 5.16: Using the MBR-based and the RANSAC-based approach for the approximation of building outlines

Chapter 6

Projection-Based 3D Building Modeling

In this Chapter a new method is proposed for generating 3D building models on different levels of detail (LOD). The proposed work flow is presented in Figure 6.1. The 3D models on different LOD follow the standard definition of the City Geography Markup Language (CityGML) described in (Kolbe et al., 2005). CityGML defines five LOD for multi-scale modeling: LOD0 – Regional model consisting of the 2.5D Digital Terrain Model (DTM), LOD1 – Building block model without roof structures, LOD2 – Building model including roof structures, LOD3 – Building model including architectural details, LOD4 – Building model including the interior.

Algorithms for producing the first three levels of the LOD are explained in this chapter. According to the above categorization, the first LOD corresponds to the digital terrain model (DTM). The non-ground regions are filtered using geodesic reconstruction to produce the DTM from LIDAR DSM (Arefi and Hahn, 2005; Arefi et al., 2007b). The LOD1 consists of a 3D representation of buildings using prismatic models, i.e., the building roof is approximated by a horizontal plane. Two techniques are implemented for the approximation of the detected building outline: hierarchical fitting of Minimum Bounding Rectangles and RANSAC-based straight line fitting and merging (Arefi et al., 2007a). For the third level of detail (LOD2), a projection-based approach is proposed resulting in a building model with roof structures. The algorithm is fast, because 2D data are analyzed instead of 3D data, i.e., lines are extracted rather than planes. The algorithm begins with extracting the building ridge lines thought to represent building parts. According to the location and orientation of each ridge line one parametric model is generated. The models of the building parts are merged to form the overall building model.
The generation of the DTM and the detection and approximation of the building outlines to form LOD0 and LOD1 have already been discussed in Chapters 4 and 5. Therefore, in this Chapter the focus lies on the generation of the LOD2 based on the analysis of the 3D points in 2D projection.

6.1 Automatic DTM Generation – LOD0

A hierarchical approach for separating the non-ground from the ground regions in LIDAR data has been presented in Chapter 4. The first-pulse (FP) or last-pulse (LP) LIDAR data are the input to a hierarchical reconstruction procedure based on different height offsets h shown in Figure 4.10. The reconstruction result is subtracted from the original data to generate the

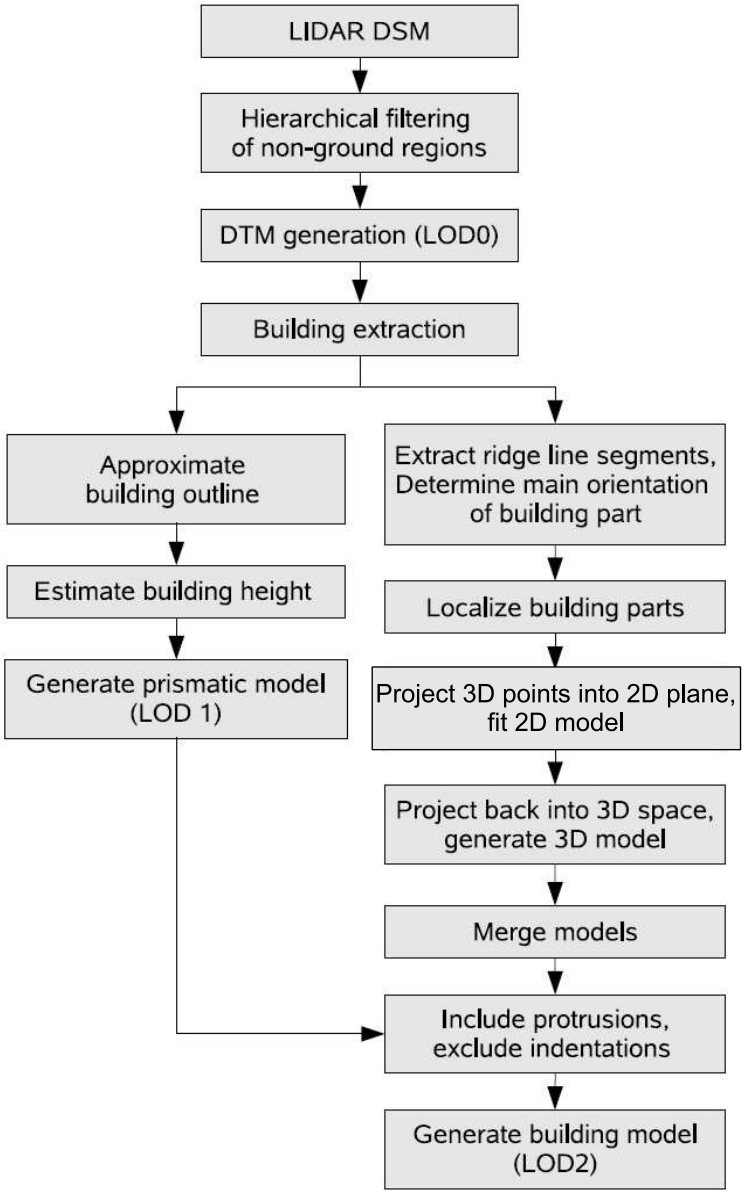
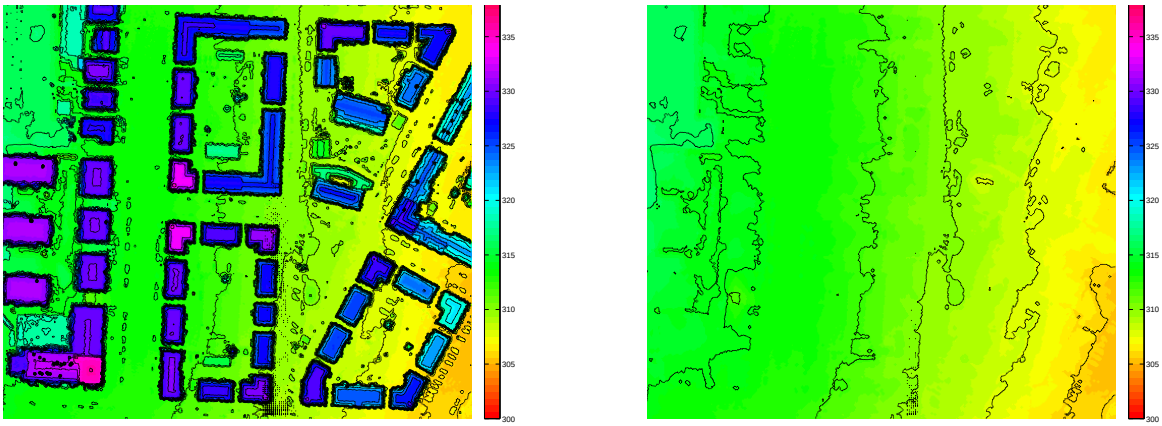


Figure 6.1: Work flow for automatic 3D building reconstruction

initial normalized Digital Surface Model ($nDSM_0$). Segmentation based on thresholding of the $nDSM_0$ and connected component analysis is utilized to generate the initial non-ground regions which are removed from the original data. Interpolation filling the gaps provides the DTM, i.e., LOD0. Figure 6.2(a) displays the LP LIDAR data with the contour lines produced from the height values superimposed. The scene consists of several buildings in a residential area with an average height difference of about 40m in about $320m \times 260m$. Figure 6.2(b) represents the corresponding DTM provided by hierarchically filtering non-ground regions using geodesic dilation.



(a) Last pulse laser data with superimposed contour lines

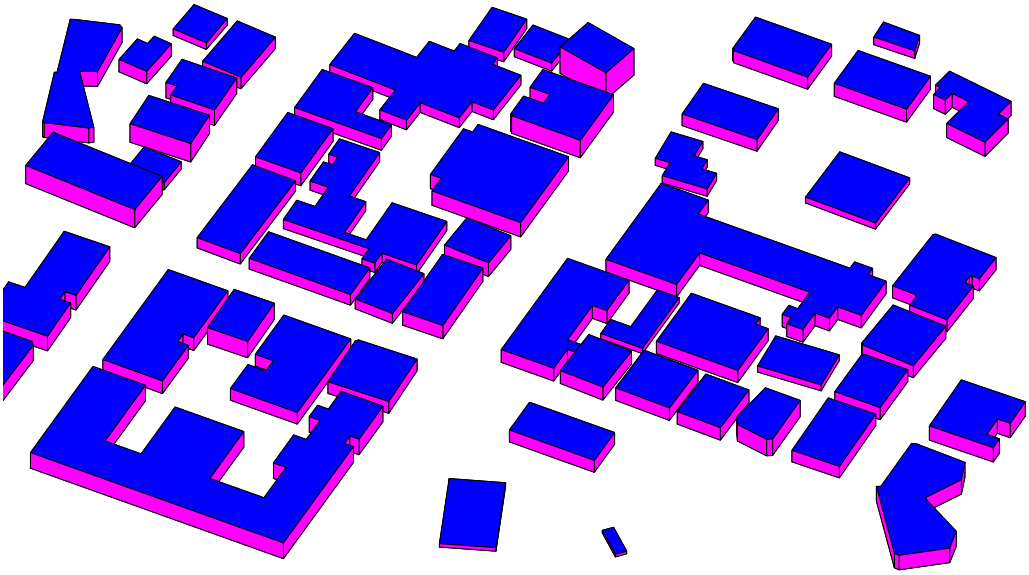
(b) DTM (LOD0) and contour lines

Figure 6.2: Generation of DTM (LOD0) by hierarchical filtering of non-ground objects

6.2 Building Outline Detection and Approximation for Generating 3D Prismatic Models – LOD1

The nDSM produced by subtracting the DTM from the original data generated in the previous section contains buildings as well as vegetation. Also other 3D objects might still be present in the data. Regions are classified rule-based using geometric and other properties of regions. The size of the regions, a vegetation index based on FP and LP range data, and variance of the surface normals have all been employed in rule-based classification (cf. Section 5.1). To model the LOD1, the extracted building outline is simplified to a polygon usually including only a few significant points, particularly corners. For this purpose, two methods are employed: fitting a rectilinear polygon by iterative adaptation of minimum bounding rectangles (MBR) and straight line fitting and merging based on RANSAC (cf. Section 5.2). To generate the 3D model from 2D polygons, the z-component is extracted from the DSM. A representative height is found by averaging the heights of the LIDAR points inside the boundary polygon. Next, the polygons representing the walls and the floor of each building are formed. All 3D polygons are put on top of the DTM to create the LOD1 representation (cf. Figure 6.4).
Figure 6.3(a) shows the original LP data in a 3D view. Building outlines are approximated by means of the combined MBR-based and RANSAC-based approach explained in Section 5.2.3. After estimating the average heights of the buildings, the 3D prismatic model is formed (cf. Figure 6.3(b)). The prismatic models are superimposed on the original LP data (cf. Figure 6.4(a)) and put on top of the DTM (cf. Figure 6.4(b)), respectively, to form LOD1.

(a) Range image – 3D

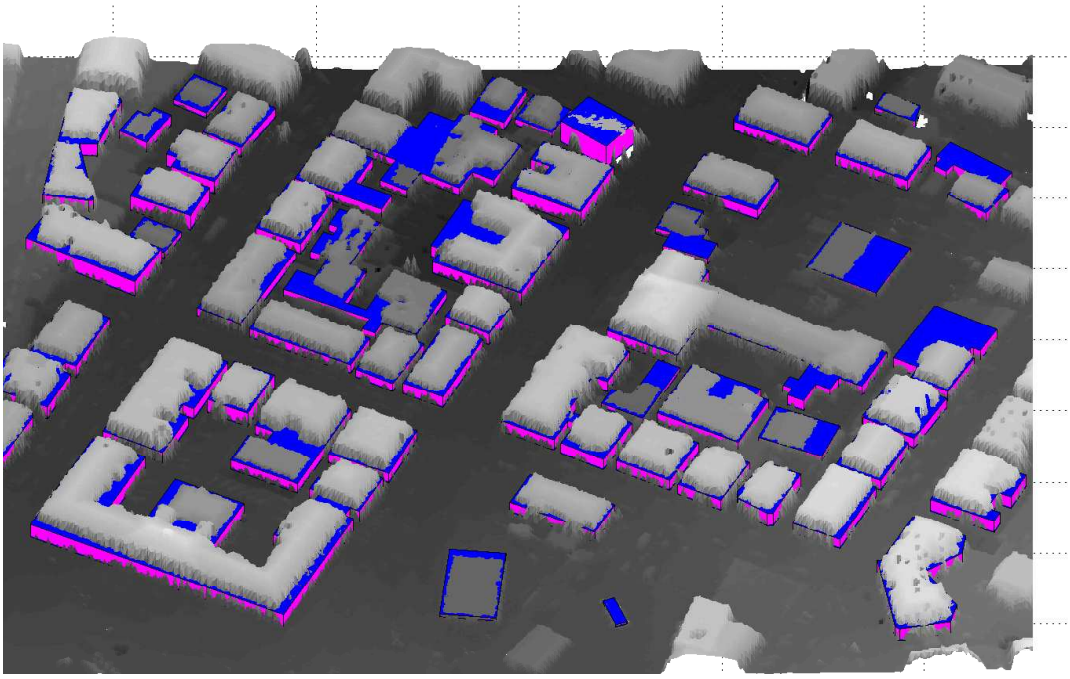


(b) 3D prismatic model generated using the combined MBR- and RANSAC-based approximation methods (cf. Section 5.2.3)

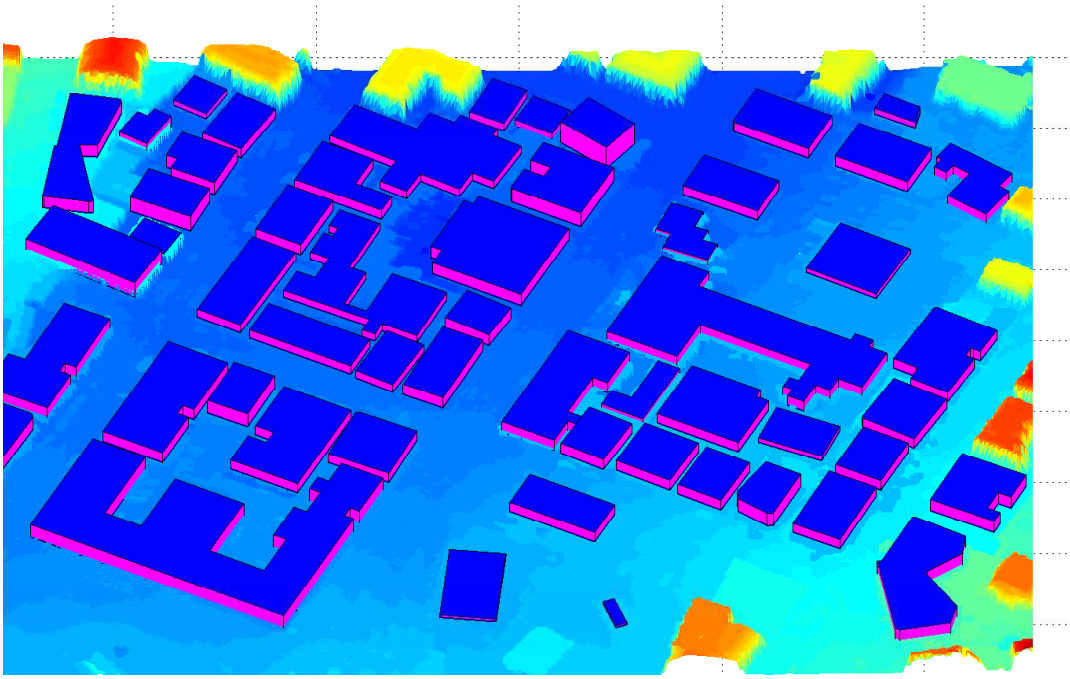
Figure 6.3: Approximation of last pulse (LP) data by 3D prismatic buildings

6.3 A Novel Approach for Building Reconstruction Using Projection Based Analysis of 3D Points – LOD2

Projection-based building reconstruction exploits that geodesic image reconstruction with a very small height difference h (cf. Figure 4.9(top)) captures ridge points and roof outlines very



(a) 3D prismatic model superimposed on original LP data



(b) 3D prismatic model put on top of DTM creates the LOD1 representation

Figure 6.4: 3D prismatic model – LOD1

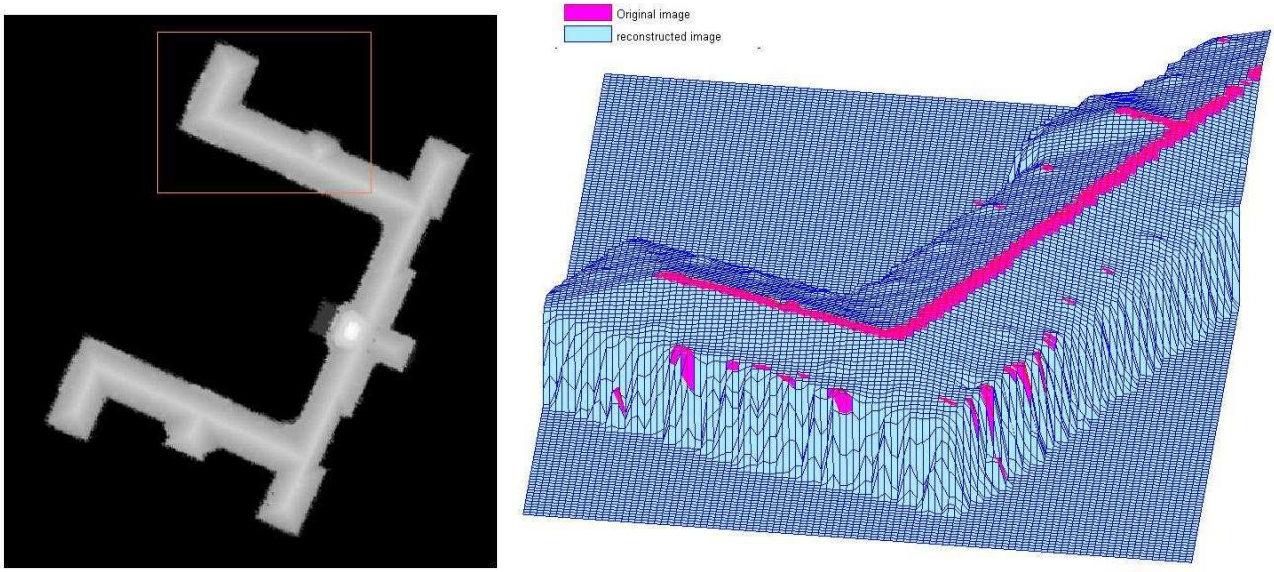
reliably. This allows to deducing the main orientation of buildings or building parts as well as a corresponding buffer zone (cf. Figure 6.7(a)). Next, a cuboid which covers the building or building part is extracted. The main orientation is used to define a 3D to 2D projection. All 3D laser points included in the cuboid are projected onto a 2D projection plane, which is one of the planes of the cuboid. It results in point accumulations in the 2D projection. The accumulations often correspond to the main building shape in terms of a profile representing the roof and the vertical walls. Here, only a limited number of roof models is taken into account, particularly flat, hipped, and gable roof.

Figure 6.7(b) shows an example with a gable roof for a building part. Robust line fitting approximates the profile by straight line segments from which a polygon consisting of the roof and the vertical walls is derived. This automatically eliminates all details of the shape the building or building part. By extruding the extracted 2D information to 3D along the normal to the projection plane, a 3D model of the building or building part is determined. The 3D model of the whole building is obtained by intersecting the models of its parts. The result is considered to be the LOD2 representation. A possible refinement to greater detail follows the same idea, but instead of analyzing all data points inside the cuboid in one projection plane, a sequence of planes is used to accumulate parts of the points of the cuboid.

6.3.1 Ridge Line Extraction and Main Orientation Determination

The ridge line extraction begins with image reconstruction by geodesic morphology extracting the pixels of the highest part of the building segment. A small height offset value, e.g., $0.2m$ is chosen for this purpose. As outcome all pixels that belong to local peaks and their neighborhood are detected as shown in Figure 6.5(b). For flat roofs the detected pixels represent the complete roof region. The region segments obtained by labeling connected components are classified into flat roof and ridge points using Gaussian and mean curvature and surface normal as features. As shown in Figure 2.6 the flat surface indicates by $H = 0$ and $K = 0$ (cf. Figure (b)) and ridge type surfaces can be identified by $H < 0$ and $K = 0$ (cf. Figure (e)), $H < 0$ and $K > 0$ (cf. Figure (a)), $H = 0$ and $K < 0$ (cf. Figure (d)) as well as $H < 0$ and $K < 0$ (cf. Figure (f)). Additionally, as mentioned, surface normal criterion is used to refine the surfaces classified by curvature values.

The number of extracted points in this step depends on the selected offset value and the inclination of the roof faces. Some other regional maxima are also detected in this step (cf. Figure 6.5(b)).



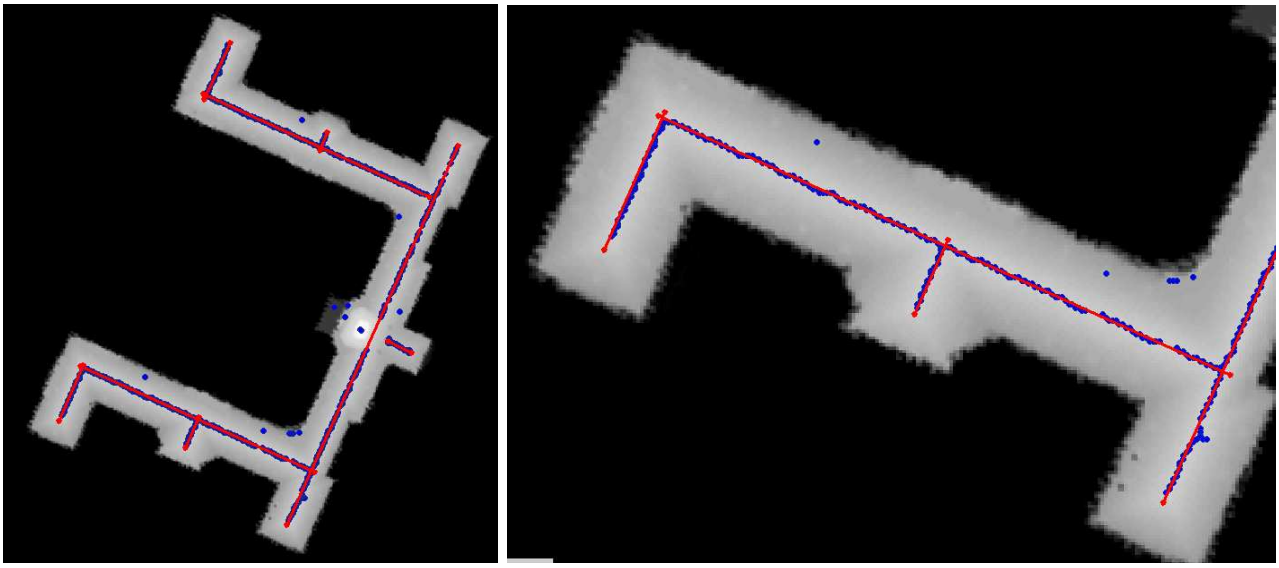
(a) Range data

(b) Roof top pixels; Differences between original and reconstructed data are represented by red points.

Figure 6.5: Determination of the roof top

Next, straight line segments are extracted from the ridge points with RANSAC. The orientations

of the ridge line segments are calculated and verified by comparison to the main orientation of the building, since in most cases the ridge lines are parallel or perpendicular to the building edges. If the deviation angle ξ between the ridge line and the main orientation is less than an empirically determined value of $\pm 5°$, the ridge line is rotated around its center of gravity by ξ. The orientation for building parts with flat roofs is calculated based on the MBR rectangle for the roof outline. Figure 6.6 shows the points classified as ridge points and the RANSAC lines superimposed on the original LIDAR image. Ridge points shown in blue in this figure are outliers of the RANSAC process or lines which are not approved because not enough inliers were found.



(a) Overall view (b) Detail

Figure 6.6: Points classified as ridge points (blue) and lines fitted by RANSAC (red) superimposed on LIDAR data

6.3.2 Localization of Building Parts

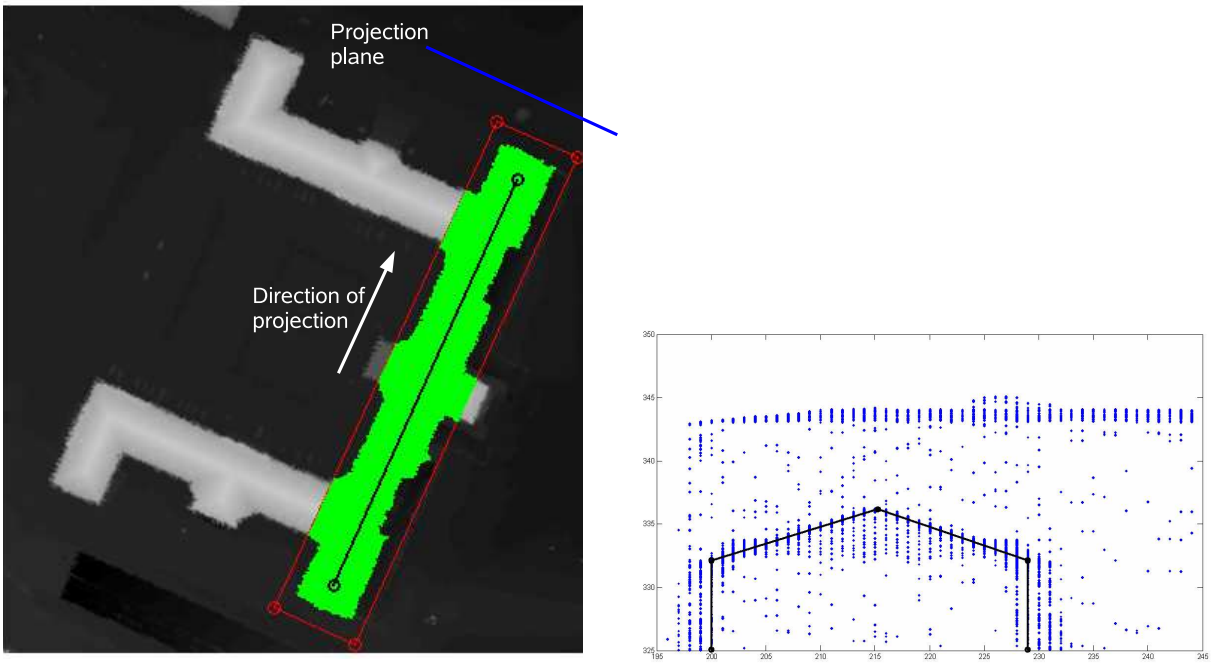
For a rectangle parallel (or perpendicular) to the main orientation, the points located inside it are extracted using the point-in-polygon algorithm. This step is necessary for buildings containing more than one part. A rectangle parallel to the main orientation, i.e., parallel to the ridge line, is created. The rectangle is defined around the ridge line with equal distances of the edges to the ridge line. The limits of the rectangle are selected in a way that the pixels belonging to the building part are all included. The orthogonal distance from the ridge to the rectangle side is calculated based on the length of the ridge (cf. Figure 6.7(a)), i.e., for a long ridge line a bigger rectangle area is considered.

In Figure 6.7(a), the rectangle is displayed in red and the localized points are shown by green dots. The direction of the projection, which is equal to the orientation of selected ridge line (black line), is given by the white arrow.

6.3.3 Projection from 3D into 2D and 2D Model Fit

The localized points from the previous Section are projected on a vertical plane orthogonal to the projection direction. According to the type of the roof which has been determined by

classification is Section 6.3.1 above, a 2D model is fitted to the projected points. For flat roofs, two vertical lines and a horizontal line connected to the top of them is adapted. For roofs with a ridge line, a model consisting of two vertical lines as well as two sloped lines intersecting at the ridge are fitted. Figure 6.7(b) shows the projected points in blue and the 2D model fitted to the data set in black.



(a) Localization of building part (green region inside buffer zone – red rectangle)

(b) Fit of 2D model (black lines) to projected data (blue points)

Figure 6.7: Projection-based modeling

The procedure to extract the 2D model from the projected points works as follows:

- The points located higher than the ridge line are not considered for 2D modeling and therefore, they are eliminated first. These points (cf. Figure 6.7(b)) usually belong to the neighboring building parts.

- To detect the two sloped lines, the points located below the ridge points are analyzed. Two lines are detected on the two sides of the ridge points which have a maximum number of supporting points. The ridge point is the common intersection point for these two lines.

- Two vertical lines corresponding to the walls on the two sides of the ridge point are detected next. As for the sloped lines, they should also have a maximum number of supporting points in their buffer zone. In this thesis the buffer zone is defined by a small distance of $0.25m$ on both sides of each tested line. Since at the eaves the quality of the laser data is possibly strongly affected by the employed interpolation method and the points are more smoothed, a robust straight line detection such as RANSAC needs to be parameterized in a way that it only extracts vertical lines.

  In this step, the number of sufficient laser points on the wall makes a significant effect on the quality of extracted wall lines. For the building parts which contains few number of laser points on their walls the quality of extracted wall in this step is poor.

- The intersection between the vertical and sloped lines defines the location of the eave points.

6.3.4 From 2D to 3D – LOD2

The 2D model is converted back to 3D by extruding it orthogonally to the projection plane. The 3D model consists of four walls plus one to four roof planes: a horizontal plane in case of a flat roof, two inclined planes in addition to two vertical triangular planes for a gable roof, and four inclined planes for a hipped roof. In the previous step the 2D model of the building part which represents the building model from front view, i.e., parallel to the ridge direction has been detected. In this Section the extension of the building part in the other direction (from the side view), i.e., orthogonal to ridge orientation is defined. The following steps complete the 3D model of the building part:

- Classify the roofs which contain ridge lines: The procedure first analyzes the points located in a neighborhood beyond the end points of each ridge line. The height at an empirical distance of about $1m$ beyond the ridge ends are used to classify the roofs into the following types:

  1. Gable roof: There exists a big negative height change (jump-down), e.g., $\delta h < -5m$.
  2. One-sided gable roof: There exists a big negative height change on one side and a big positive height change (jump-up), e.g., $\delta h > 5m$, on the other direction. This happens when the building part in one side connects is on a lower height level than a neighboring part.
  3. Two-sided gable roof: On both sides the height changes are positive. It happens when the building part is located between two other parts on a lower height level.
  4. Hip roof: The heights change smoothly with negative values, e.g., $0 > \delta h > -1m$.
  5. Gable-Hip roof: On both sides the height changes are negative, but on one side with a large value, i.e., $\delta h < -5m$, and on the other side a small value, i.e., $0 > \delta h > -1m$.

- If the building part consists of jump-down or jump-up points, the building model is extended up to these points. In other words, if the 2D model is generated by analyzing the points in the XZ plane, the Y elements of the walls on one side (or on both sides in case of roof types number 1, 2, and 3) get the Y value of the ridge point on that side.

- If according to the classification the building contains a hip on one or both sides, i.e., cases number 4 and 5. In case of hip roof, a straight line detection method using a fixed point, i.e., the ridge point and the point in the neighborhood, which produces the line with the maximum number of supporting points, is employed. This inclined line represents the hip roof from side view. In this case, the Y element of the wall is defined by the same procedure as the 2D model. In case of gable roof a vertical line from the ridge point to the ground is formed as wall in side view.

- The height of the ground is the only remaining unknown value to form the 3D model of the building part. Here, the minimum height value of the points corresponding to the building part is defined as ground height.

- Finally, the polygons for the faces, i.e., roof, walls, and floor, are shaped using the coordinates of the nodes and the ground height.

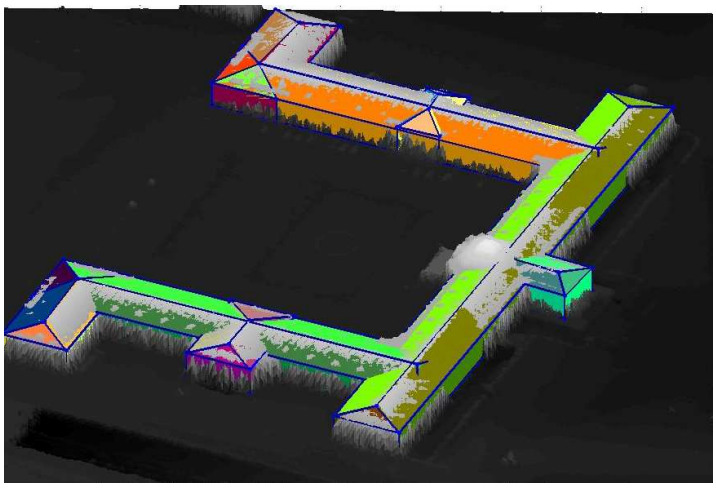


Figure 6.8: 3D building model for new Stuttgart castle. Colored faces of the model are visible only if they are above the gray data points.

After reconstructing 3D models for all building parts, they are merged to form the overall 3D model of the building. Figure 6.8 displays a building model produced by merging eight building parts. The eight ridge lines lead to eight parametric building models with hipped roofs. After combining the models, an overall 3D model is provided. For nodes of the building parts which have to be merged because they represent, e.g., the same corner of the building, the average value is determined.

Figure 6.8 shows one building part which is not modeled directly above the entrance of the Stuttgart new castle. A proper model for this area would be a dome. Since the model is complex and is not included in our roof classification it could not be modeled.

As shown in Figure 6.11, intersections might need to be refined for not rectangular building parts. To refine the two inconsistent nodes, the distances between the original LIDAR data points and each eave line are calculated. For each eave line closest points, i.e., having a distance less than a certain threshold, are extracted. Using these supporting points the eave lines are extended (or shortened) from both sides to the last point. For nodes generated from more than one vertex the average is used.

2D information about protrusions and indentations of the building boundaries can be extracted from LOD1. The nodes of the protrusions and indentations are determined from the approximated polygon and the corresponding planes on the roof (flat or inclined ones) are adapted. Figure 6.9 displays an approximation result for the new Stuttgart castle superimposed on the original data as red polygon. The reconstructed building model is given in blue. The figure shows that the 2D outline of the approximation and the overall 3D outline almost fit together. One protrusion and two indentations exist which have to be included or excluded from the model. In this example the protrusion is a low building part with flat roof modeled using a cuboid. Both protrusion and indentation regions are excluded using information from the approximated outline, i.e., a prismatic model.

By comparison between the overall boundaries of the 3D building models and the approximation result for the building block, the corresponding indentation and protrusion regions can be detected. The overall area of the approximation result for the building block is subtracted from the corresponding area for the parametric models. This is done by subtraction a binary image which highlights the internal pixels of the approximation result from the corresponding binary

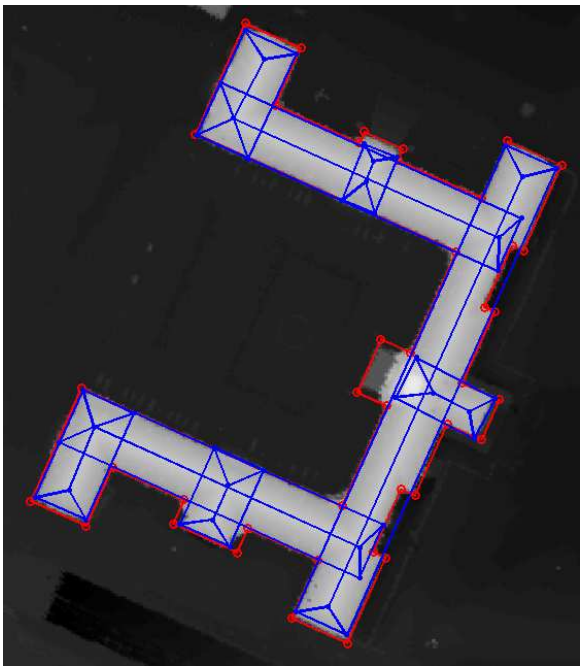


Figure 6.9: Approximation of the building outline (red polygon) and the reconstruction result (blue polygons) superimposed on the original data

image of the parametric models. The positive pixels belong to protrusions and the negative pixels relate to indentations.

The regions provided by connected components labeling are analyzed and small regions are eliminated. An MBR is adapted to each remaining region after smoothing the regions using morphological closing. For protrusion, a cuboid-shaped 3D model is generated and the average of the heights of the internal points is used as building height. Although, this does not mean that the protrusion parts have always flat roof, but since their corresponding roof types cannot be distinguished by the proposed algorithm in this thesis, a simple cuboid shaped model is fitted to the points. The corresponding MBR polygon nodes of indentation and protrusion regions are included in the overall 3D model. Finally, the inclinations of the building roofs are adapted after including the indentation nodes (cf. Figure 6.10).

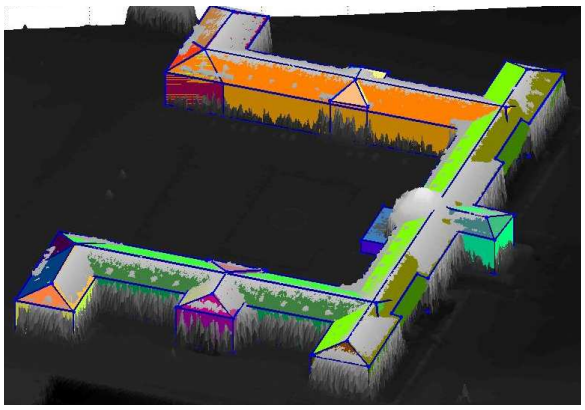
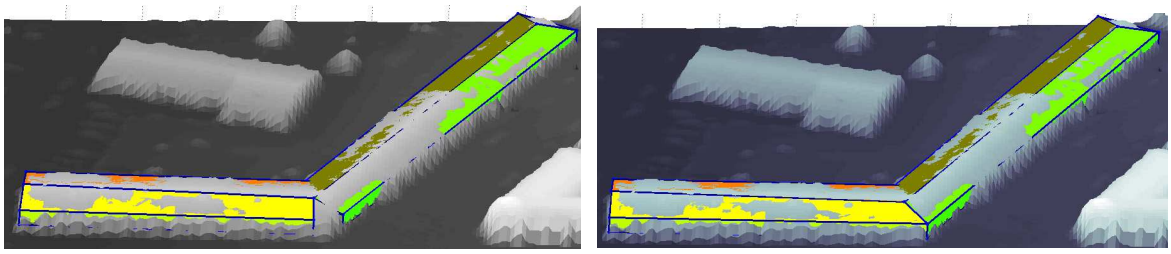


Figure 6.10: 3D model extended by a protrusion and two indentations (colors cf. Figure 6.8).



(a) Before adjusting the nodes (b) After adjusting the nodes

Figure 6.11: An example for projection-based building modeling (colors cf. Figure 6.8) with building parts intersecting at a non-orthogonal angle.

Chapter 7

Experimental Investigations with Stuttgart Test Data

7.1 Test Data

The algorithms proposed in this thesis were tested and evaluated with data from the city of Stuttgart, Germany. The company TopScan GmbH (TopScan, 2007), Rheine, Germany kindly provided LIDAR range and intensity data covering the central part of Stuttgart. The data are recorded in combined first (FP) and last pulse (LP) mode. Airborne LIDAR data were acquired in April 2006 by an Airborne Laser Terrain Mapper (ALTM) laser scanner, *ALTM* 2050, with a measurement density of about 4.8 points per m^2 on average. The specification of the LIDAR data used in this thesis is presented in Table 7.1.

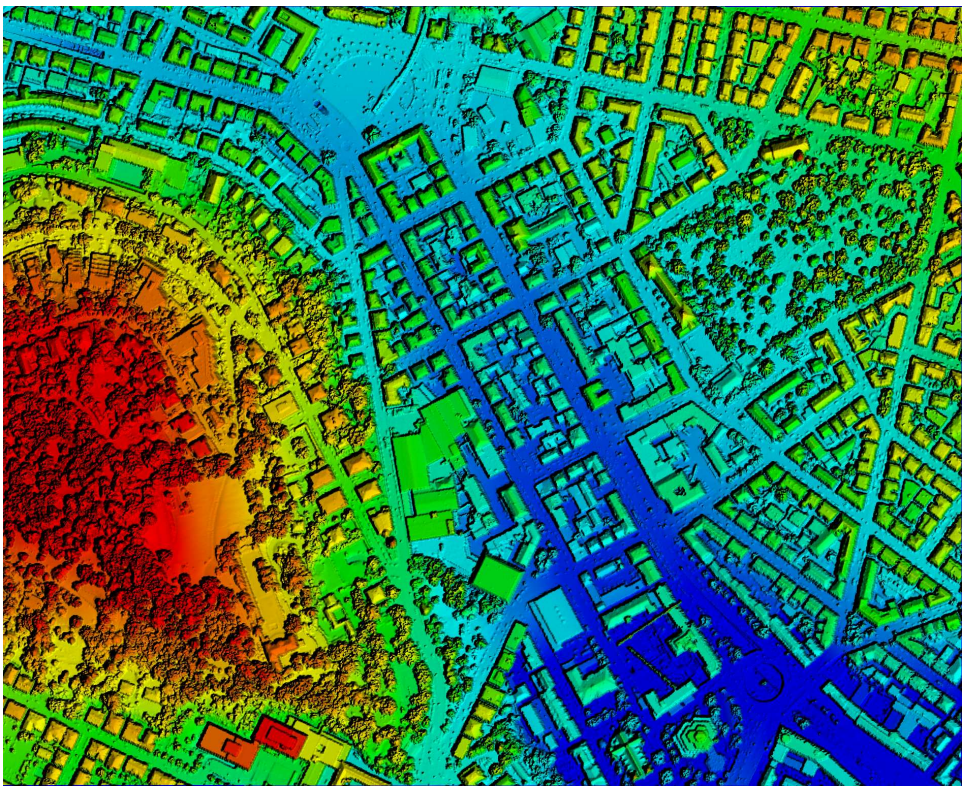
Specification	**Data**
Capture period	April 2006
Flight altitude	1100 meter
Pulse frequency	50.000 Hz
Wavelength of the Laser	1064 nm (near infrared)
Measurement density	$\approx$ 4.8 points per m^2
Swath width	$\approx$ 600 meter
Instrument	ALTM 2050 laser scanner

Table 7.1: Airborne LIDAR data specification

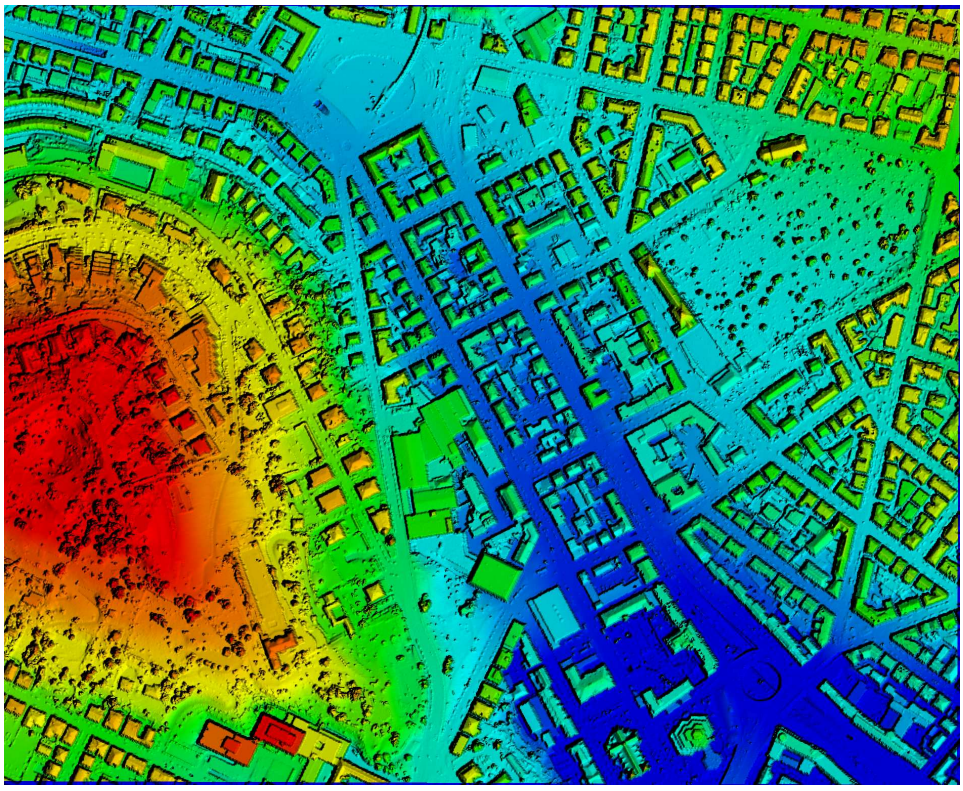
A regularly spaced elevation grid is derived by means of spatial interpolation of the raw 3D points. A $0.5m$ lattice spacing is chosen for the elevation grid. A sample of the test data visualized as color image is represented in Figure 7.1. It represents a $1km^2$ hilly residential area with a maximum height difference of about $200m$. Both FP and LP data are visualized. The area contains different $3D$ objects on top of the undulated terrain such as very dense vegetation regions as well as spacious buildings.

7.2 Experimental Results

In this Section results of each processing step are given and analyzed. The processing begins with filtering the LIDAR data based on geodesic dilation to separate non-ground and ground



(a) First pulse DSM



(b) Last pulse DSM

Figure 7.1: LIDAR DSM on regular grid format with $0.5m$ grid size

regions, resulting in DTM and nDSM (cf. Section 7.2.1). Next, rule-based classification leads to buildings, vegetation, and other classes which are analyzed in Section 7.2.2 together with the building outlines approximated and regularized based on MBR and RANSAC. 3D modeling of the ground pixels and reconstruction of the buildings according to the first three levels of detail (LOD) defined by CityGML is the final processing step. The result for each LOD is presented and their accuracy is analyzed in Section 7.2.3.

7.2.1 Separation of Non-ground and Ground Regions – Generation of DTM and nDSM

The separation of ground and non-ground points is based on the LP elevation grid. After removing all non-ground points from the LIDAR data and filling the gaps using appropriate interpolation technique, the DTM is obtained.

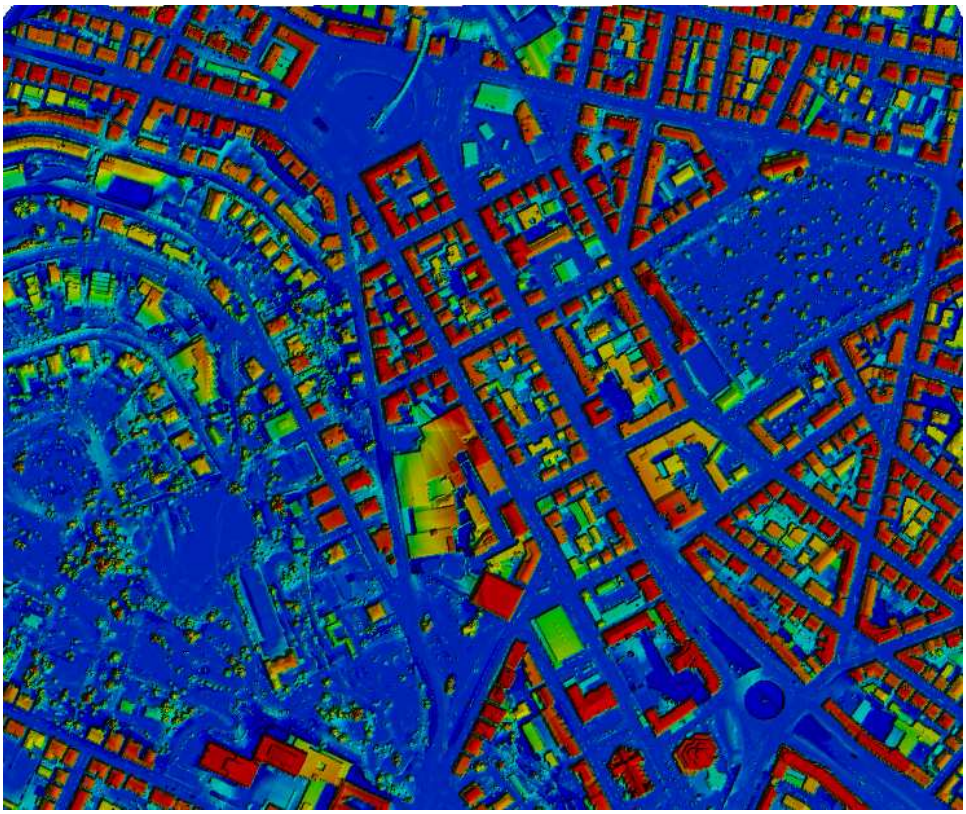


Figure 7.2: Resulting nDSM

Figure 7.2 illustrates the nDSM generated by the reconstruction method. It is produced by subtracting the DTM from the original LP data (cf. Section 4.3).

The non-ground regions are iteratively filtered using different sequentially generated *mask* and *marker* images. In each step an initial nDSM (nDSM0) is generated by subtracting the reconstructed image from the *mask* image. The detection uses simple thresholding on the nDSM0. Foreground regions are created and evaluated based on Local Range Variation (LRV) (cf. Figure 7.6(b)).

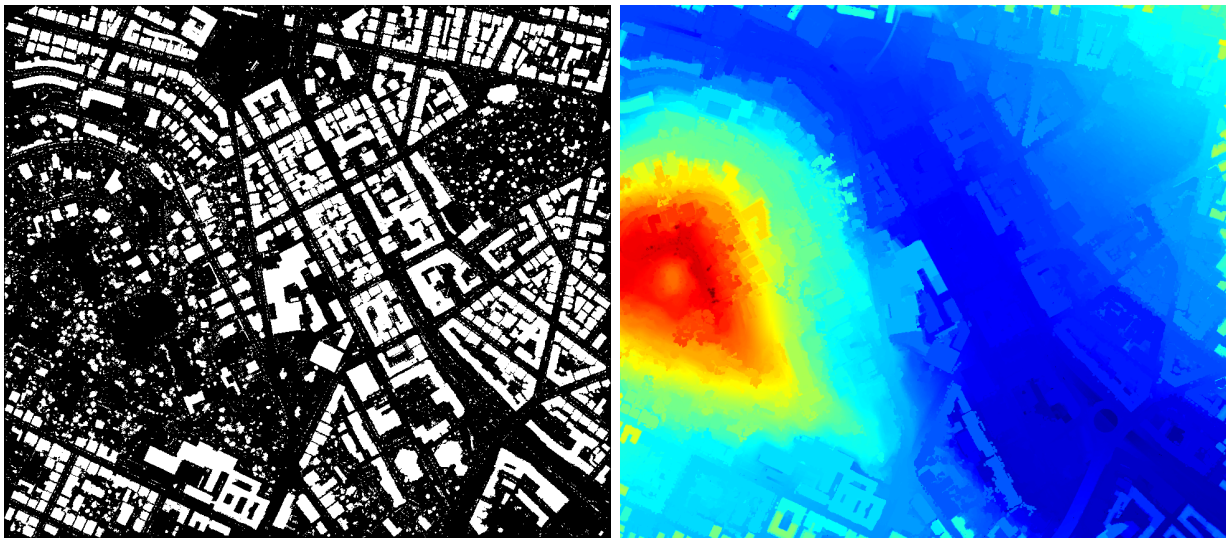
Above the terrain regions are detected iteratively with the number of iterations depending on the maximum value of height change extracted from the LRV image. To increase the processing speed, geodesic dilation begins with the maximum offset value which is equal to the maximum

value of the LRV. This is because the smaller non-ground regions are usually included in the larger regions. I.e., if a big region such as complex building is extracted using a big offset value, all the internal parts are extracted as well. Consequently, the smaller internal parts do not need to be extracted and evaluated any more.

The *marker* image is produced based on the offset value $h = max(LRV)$ except for boundary pixels where $marker = mask$. The result after image reconstruction with offset $h = max(LRV)$ is subtracted from the original range data, i.e., LP, to generate the $nDSM0$. A threshold value of about $30cm$ is used to avoid highlighting very low objects. The potential non-ground regions are produced and their boundary pixels are analyzed to distinguish 3D non-ground regions.

This procedure repeated for all other offset values and all the non-ground regions in each step are integrated in one binary image. Figure 7.3(a) represents the result of non-ground regions extracted from LP data. Subsequently the non-ground regions are eliminated from the original LP image to make not-filled DTM. Finally the gaps are filled by means of IDW interpolation (cf. Section 4.3) to generate DTM or filled DTM (cf. Figure 7.3(b)).

Figure 7.3(a) shows that almost all of the above ground objects such as buildings and trees are correctly extracted. Comparing this Figure and original LP (cf. Figure 7.1(b)) shows that there are still some 3D objects in particular those are locating on the hill which are not correctly extracted. Figure 7.4 enlarged one of the buildings that are not appropriately extracted.



(a) None-ground regions

(b) DTM generated after eliminating the non-ground pixels and filling them with IDW interpolation

Figure 7.3: Non-ground regions are eliminated from LP and the gaps are interpolated to form DTM. The produced DTM is not complete and some building parts are still remained.

Geodesic reconstruction filters objects which are higher than their surroundings. I.e., all pixels on the boundary of an object must have a bigger value than the pixels outside next to the object. Objects to be filtered (such as buildings and trees) are often higher than their neighborhood. Yet, in the suburban hilly regions it happens occasionally that spacious buildings are situated on steep terrain. One such building is presented in Figure 7.4 in the center of the test data (cf. Figure 7.1(b)).

Here, the road next to the building on the left side has a larger height ($327.0\,m$) than a part of the building on the right side ($322.6\,m$). This means that not all parts of the building are higher than its neighboring pixels. In this case, the morphological reconstruction is not able to

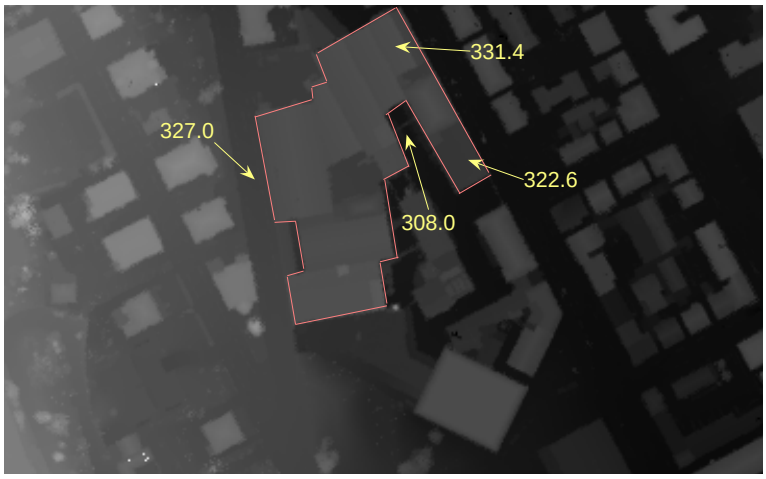


Figure 7.4: Special case spacious building block: The height value on the roof on the right side (322.6 m) is less than the height of the road next to it on the left side (327.0 m)

filter these objects in the procedure explained above, as can be seen in Figure 7.3(a): Only the parts of the building block which are higher than the neighboring areas are filtered.

In order to efficiently extract such non-ground objects the algorithm needs to be improved. For that, after each iteration, the non-ground regions are replaced by a constant value in the *mask* image. A big constant value with minus sign acts as below ground outlier and has theoretically no effect in image reconstruction procedure. The image reconstruction by dilation only modifies the foreground objects and below ground objects remain intact. The new *mask* image with a constant value replaced in the non-ground regions detected from previous step used for the processing in the next step.

Figure 7.5 represents the processing steps for extracting the spacious building discussed above (cf. Figure 7.4). The process begins with an offset value of $h = 21m$ which is measured by the maximum value of LRV. Figure 7.5(a) shows the original image LP regarding to this building. The non-ground regions extracted in the first step are replaced by $-1000m$ (cf. Figure 7.5(b)). This image (cf. 7.5(b)) is utilized as new *mask* image for the next iteration.

After producing the new *mask* based on the first offset value, a new LRV called $LRV0$ is generated based on the new *mask*. The maximum value of the $LRV0$ is computed and the iteration values for the offset h are defined as:

$$h = max(LRV0) : -inc : 1$$

A small increment *inc* is used to make sure that all 3D objects, particularly buildings, which are situated on different height levels, will be extracted. An increment value of $1m$ has been chosen empirically.

The procedure continues until the iteration is finished. Some of intermediate results are shown in Figure 7.5 (from (c) to (f)). Figure 7.5(f) illustrates that in final result the building is completely extracted.

Figure 7.6(a) displays all detected non-ground regions which are produced from the complete scene (cf. Figure 7.1(b)). For the processing of this scene the following offset values are generated:

- for the first iteration:

$$h = max(LRV) = 21m$$

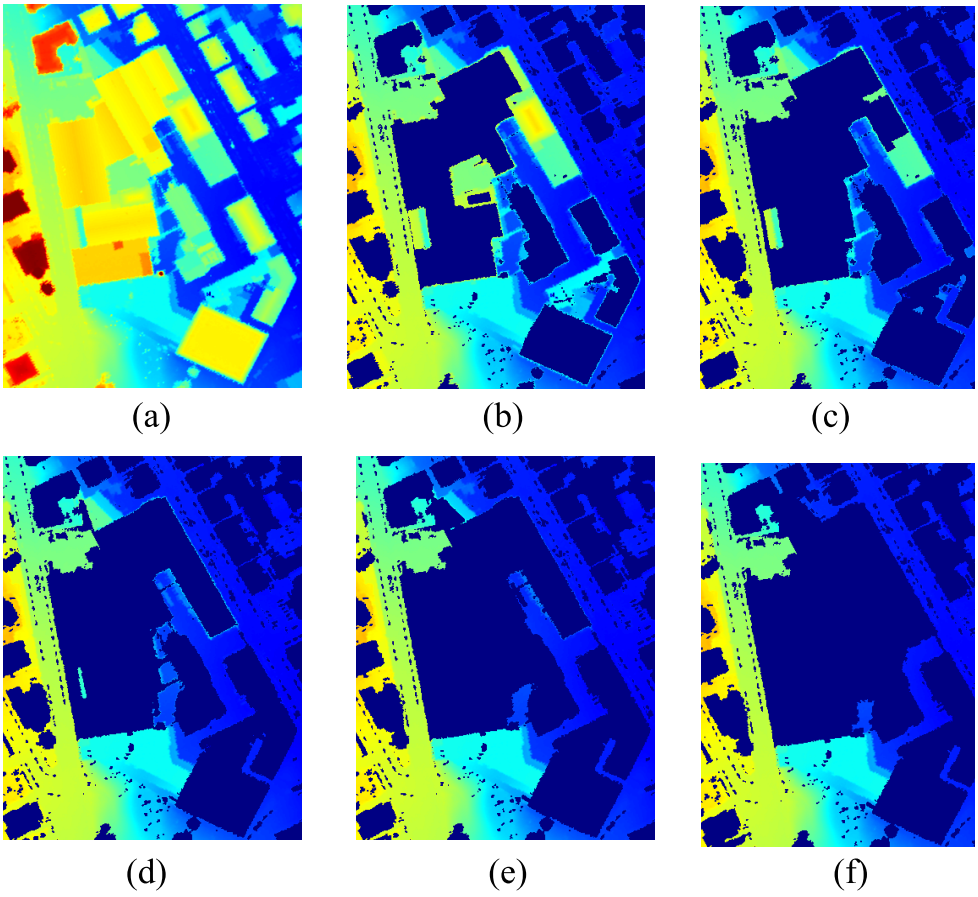


Figure 7.5: Processing steps of filtering spacious building block: (a) original LP image, (b) to (f) extracted non-ground regions (dark blue) by employing: (b) h=21m, (c) h=8.2m, (d) h=7.2m, (e) h=6.2, (f) h=1.2m as offset value for marker generation

- for the next iterations:

$$h = max(LRV0) : -1 : 1 = \{8.2m, 7.2m, 6.2m, 5.2m, 4.2m, 3.2m, 2.2m, 1.2m\}$$

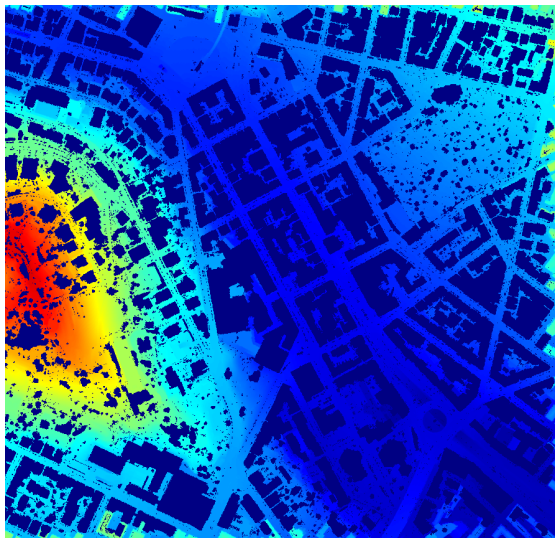
The non-ground regions are highlighted in Figure 7.6(c) with dark blue color. A DTM is generated by removing all non-ground points from the original LP data and then filling the gaps by interpolation (cf. Figure 7.7).

As can be seen in Figure 7.7 and all other figures related to geodesic morphological processing, the 3D objects located on the boundary of the image are not filtered. This is due to selecting there the same values in the *marker* image the as in the *mask* image. Therefore, if the pixel on the boundary belongs to a building roof, the geodesic dilation proceeds from this point and the building part is not filtered. Geodesic dilation only filters the foreground object if all its boundary pixels are higher than the neighboring pixels and this is not the case for objects which are not completely inside the image. Thus, a bigger area than the desired region should be processed with the size of the processing area large enough so that all 3D objects of interest are entirely inside the image.

Figure 7.7 shows that almost all obvious non-ground regions have been eliminated. These regions mainly represent buildings and vegetation areas. None of the buildings not on the

(a) None-ground regions after final iteration – in this hilly area the number of iterations is 9

(b) Local Range Variation (LRV) which highlights height jumps



(c) Non-ground regions *(dark-blue)* eliminated from original image

Figure 7.6: Non-ground regions segmented using morphological geodesic dilation and evaluated by LRV feature

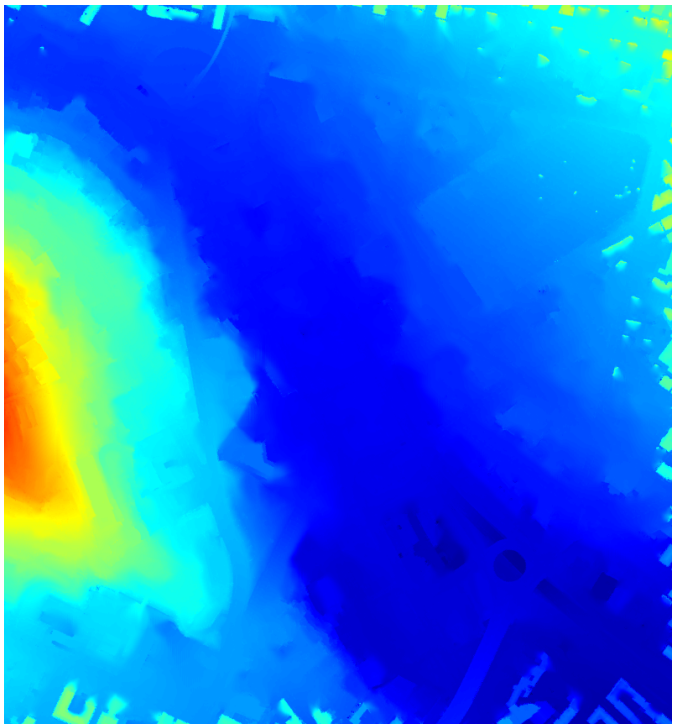


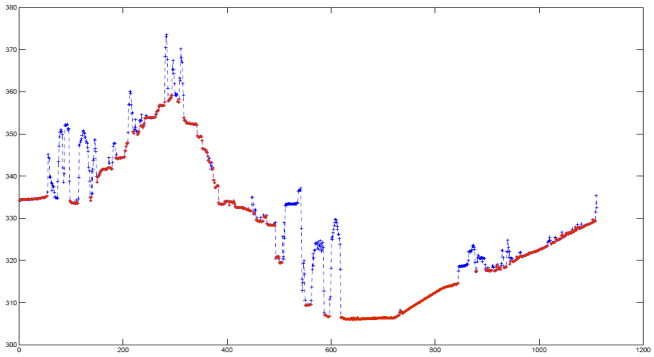
Figure 7.7: DTM after interpolation of gaps

image margin is visible any more and also the vegetation has disappeared. Shape and size of the objects are obviously irrelevant for our approach. Large buildings as well as small ones, elongated buildings as well as short ones, and high buildings as well as low ones have been properly eliminated.

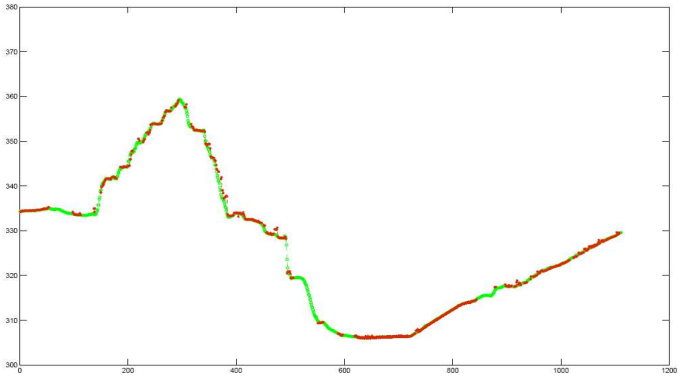
To evaluate the quality of the result in addition to visually evaluating the results, profiles in the image as well as a comparison of the derived *DTM* with the original data and ground truth data are given. The ground truth data is a DTM collected manually by the surveying agency *(Landesvermessungsamt)* of the state Baden-Württemberg in Germany in 2002. There were some changes in the area from 2002 to 2006, but the ground truth still can be used for evaluation.

Profile No. 1 (cf. Figure 7.8) ranges from the lower-left to the upper-right corner of the image and profile No. 2 (cf. Figure 7.9) from the lower-left to the lower-right corner. Height values of the original *LP (dashed blue lines)*, of ground points extracted using the proposed approach *(red points)*, of the *DTM* after filling the gaps *(black lines)*, and of the ground truth *DTM (green lines)* are compared. For a better evaluation of the quality of the results, the height values are shown in three different sub-figures. Visual comparison proves that the non-ground objects are appropriately filtered from the original dataset. Even big buildings located on hilly terrain are properly separated from the ground points.

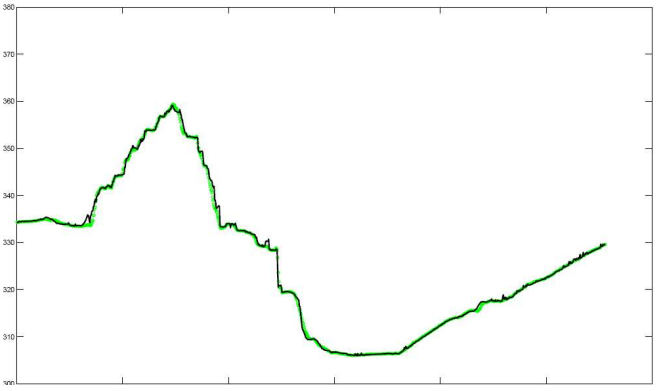
Histograms are used to quantitatively evaluate the produced DTM in comparison to ground truth. One histogram presents the height values of the ground points extracted by the proposed algorithm (cf. Figure 7.10 left) and another of the ground truth DTM (cf. Figure 7.10 middle). Additionally, for a better comparison the histogram of the difference between the two datasets (cf. Figure 7.10 right) is given. Since the proposed approach based on geodesic morphology only filters objects entirely located inside the image, there are some obvious large differences between the ground points and the ground truth at the image border. To decrease their effect on the evaluation, margins of the images are not taken into account for histogram production.



(a) Ground pixels *(red)* versus original last-pulse pixels *(blue)*

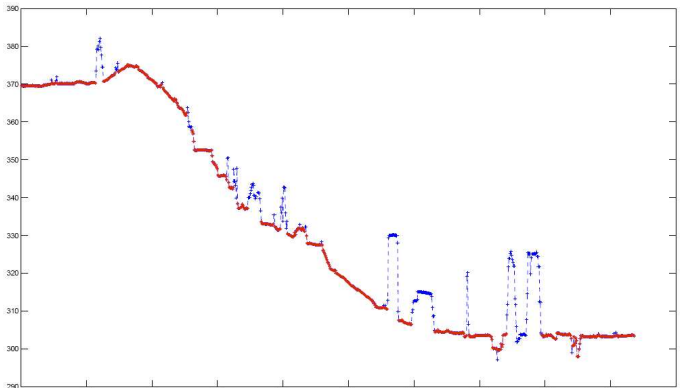


(b) Ground pixels *(red)* versus ground truth DTM *(green)*

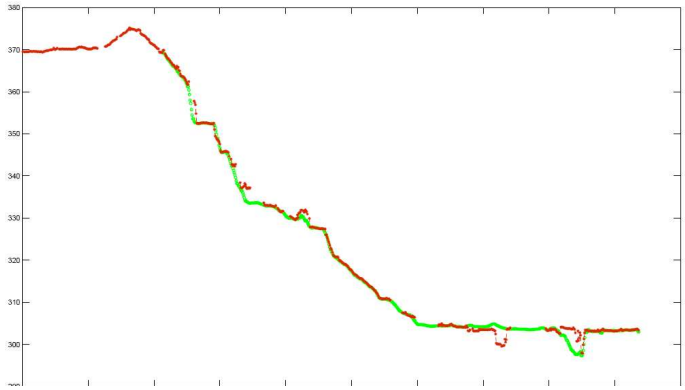


(c) DTM provided by proposed approach (black) after interpolation of gaps versus ground truth DTM (green)

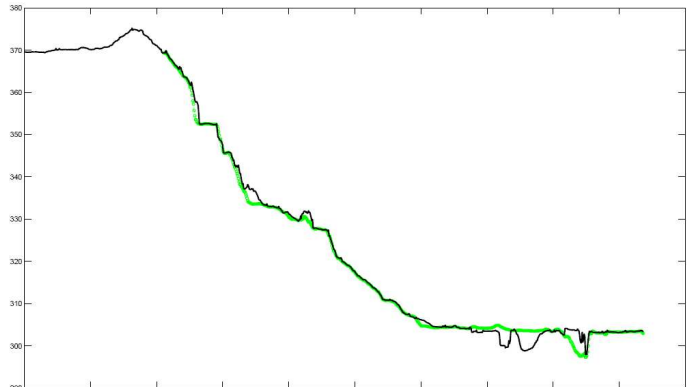
Figure 7.8: Profile No.1 – profile ranges from lower-left corner to upper-right corner of the area without margins



(a) Ground pixels *(red)* versus original LP pixels *(blue)*



(b) Ground pixels *(red)* versus ground truth DTM *(green)*



(c) DTM provided by proposed approach (black) after interpolation of gaps versus ground truth DTM (green)

Figure 7.9: Profile No.2 – profile ranges from lower-left corner to lower-right corner of the area without margins

The evaluation of the generated DTM based on profiles and histograms in comparison to corresponding ground truth proves that:

- Almost all 3D objects, i.e., buildings and vegetation, are filtered from the LP image except objects not completely situated inside the image.
- Objects which are not higher than the neighboring regions cannot be filtered by the given approach. Buildings that are partly outside the image are as one example. Horizontal bridges, i.e., bridges that cross another street at a low height are as another example. These kind of objects have large discontinuities on two sides and almost no discontinuity at the other sides.
- The histogram in Figure 7.10-right shows that the difference between the produced DTM and ground truth is in most cases around zero:
  - 86% of the pixels have an absolute difference of less than $0.5m$
  - 92% of the pixels have an absolute difference of less than $1m$
  - The standard deviation of the differences is about $0.56m$
- Spacious building blocks as well as small objects can be filtered at the same time in hilly areas without adaptation of the size of structuring element. The latter is a problem for standard filtering approaches based on basic morphological operations.
- The approach is fast and the speed of processing is mainly depends on the hilliness of the area. In regions with low height variation usually all objects can be filtered in $3-4$ iterations.

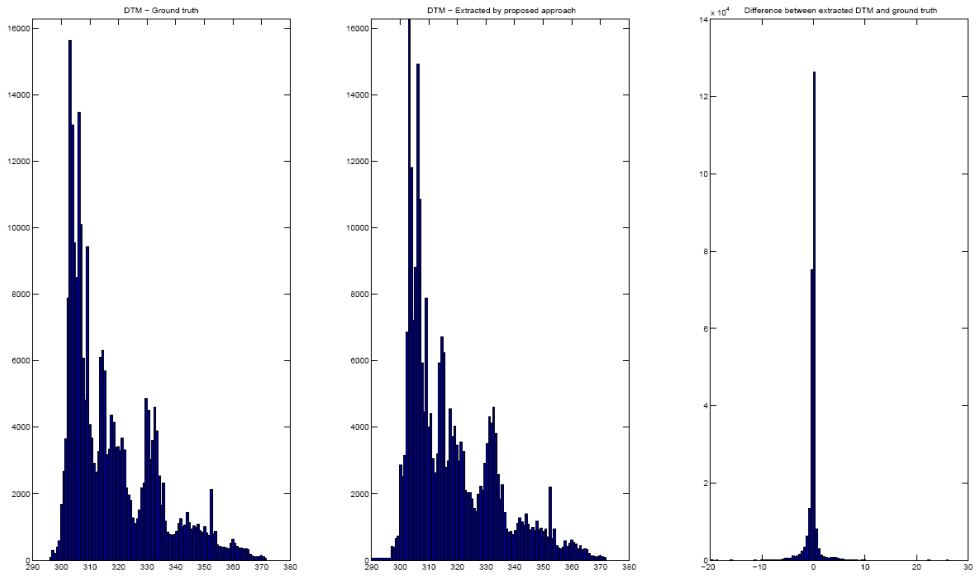


Figure 7.10: Histograms for ground pixels, ground truth DTM, and their difference.

The error or difference between the generated DTM and the ground truth seems not to be very significant if one considers that the computations are based on different observations which are

acquired at different epochs. The other point is that the DTM generation method is different and different criteria might be used to classify a region as a 3D object. This might happen mainly for low objects such as low vegetation. Finally, the interpolation technique to generate raster DSM and DTM might be different.

7.2.2 Classification of LIDAR Data and Approximation of Building Outlines

Results of building and vegetation detection as well as for the approximation of building boundaries are given and discussed in this Section. The classification of non-ground pixels is tested with a data set that contains buildings of different sizes as well as bushes, single trees and groups of trees, the latter summarized under the term vegetation in the following.

The rule-based classification of non-ground LIDAR pixels into building, vegetation, and crane or power lines has been explained in detail in Section 5.1. While generating feature descriptors as classification criteria a vegetation index VI feature is defined (cf. Figure 5.3) based on FP and LP data. The VI is segmented using thresholding. The VI values as well as the original height values of the points inside each region are analyzed to produce two main classes namely vegetation and crane (or power line) (cf. Figure 5.4). Not all vegetation pixels are classified in this step. Since the VI highlights only pixels having a height difference between FP and LP, vegetation pixels will (almost) the same height value in FP and LP are not detected in this step.

Local range variation LRV (cf. Figure 7.6(b)) is the second feature descriptor created from LP. It helps to separate the objects with large discontinuities, e.g., buildings or trees, from other objects.

To finally separate the buildings and the rest of the vegetation, i.e., the vegetation which have the same height in FP and LP, the $nDSM$ (cf. Figure 7.2) is analyzed. The $nDSM$ pixels are thresholded and initial regions are produced by connected component labeling. The LRV values of the region boundaries as well as variances of the surface normals of the heights inside each region are employed to classify the segmented regions into building, vegetation, and other objects. The vegetation regions classified in this step are combined with the results of the first step to generate the complete vegetation class.

Figure 7.11 shows the finally classified buildings and vegetation. In this figure the boundaries of the buildings and vegetation regions are superimposed on the FP data in *blue* and *red* color, respectively.

To quantitatively investigate the quality of the extraction, ground truth data are manually digitized using the LP (cf. Figure 7.12). The focus is laid on the buildings and the quality parameters correctness and completeness are determined as defined by Heipke et al. (1997). To do so, the building extraction result is labeled as true positive *(Tp)*, false negative *(Fn)*, or false positive *(Fp)* on a pixel by pixel basis. Then, completeness represents the ratio of correctly extracted pixels with respect to the total number of pixels of each class from the ground truth and is defined as:

$$completeness = \frac{Tp}{Tp + Fn}$$

$$completeness \in [0; 1]$$

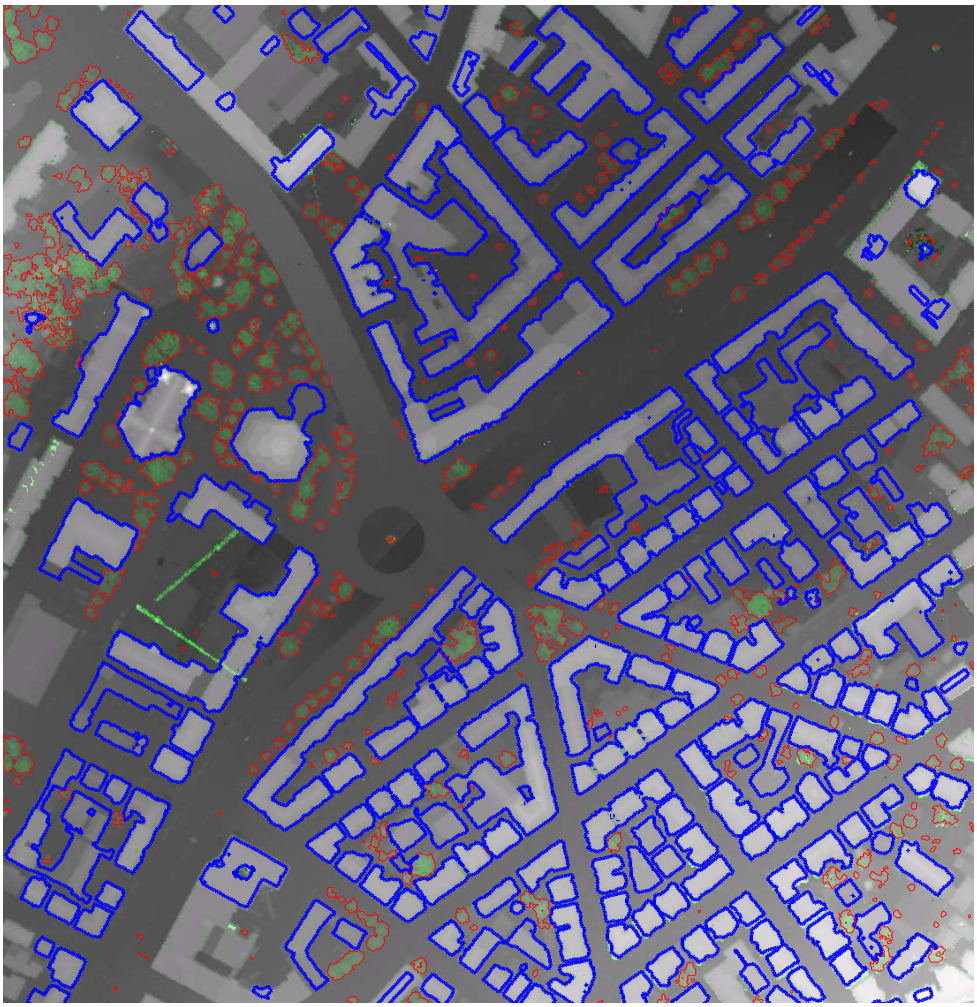


Figure 7.11: Classified building *(blue polygons)* and vegetation *(red polygons)* regions are superimposed on the gray scale LIDAR image that the pixels belonging to the vegetation index are colorized by green.

Correctness indicates the ratio of correctly extracted pixels with respect to the total number of extracted pixels of each class and is defined as:

$$correctness = \frac{Tp}{Tp + Fp}$$
$$correctness \in [0; 1]$$

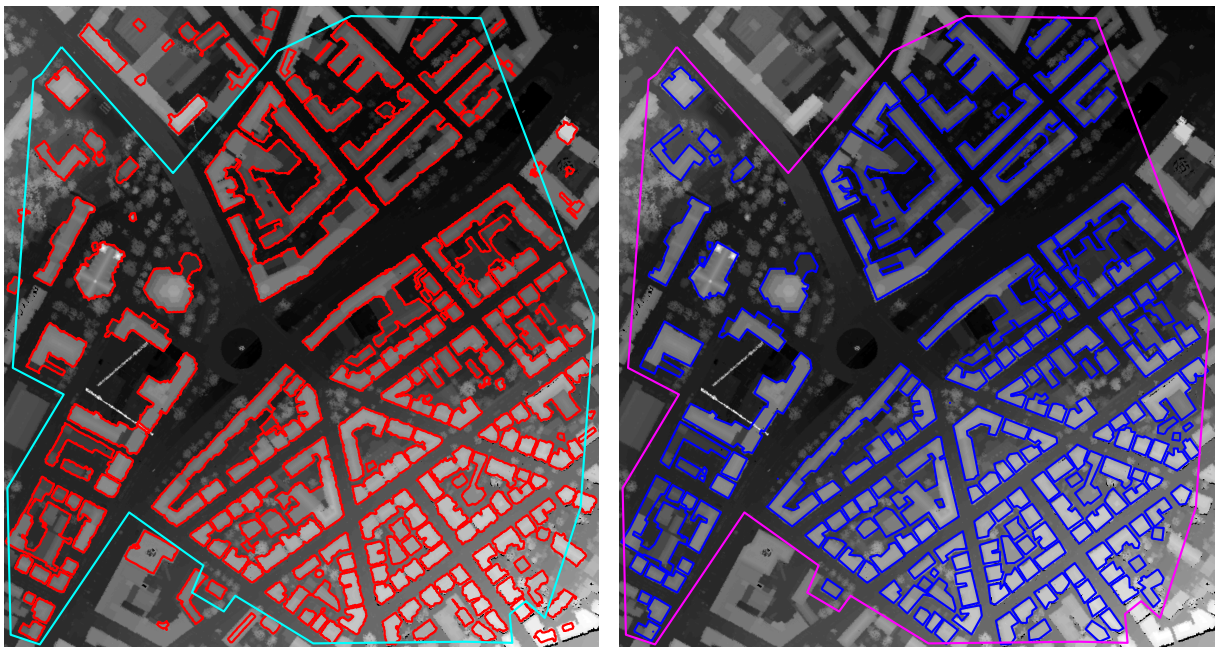
and finally, the Quality parameter is defined as:

$$quality = \frac{Tp}{Tp + Fp + Fn}$$
$$quality \in [0; 1]$$

The optimum value for all defined parameters is 1. To calculate the parameters, an image labeled according to the extracted regions and the ground truth is provided (cf. Figure 7.13).

Accordingly, the following values have been computed:

$$completeness = 0.93$$



(a) Building polygons *(red)* produced by classification and study area *(cyan)* overlaid on FP data

(b) Manually digitized building ground truth *(blue)* and study area (magenta)

Figure 7.12: Extracted building polygons, corresponding ground truth, and limit of the study area guaranteeing that all buildings are entirely situated inside the image

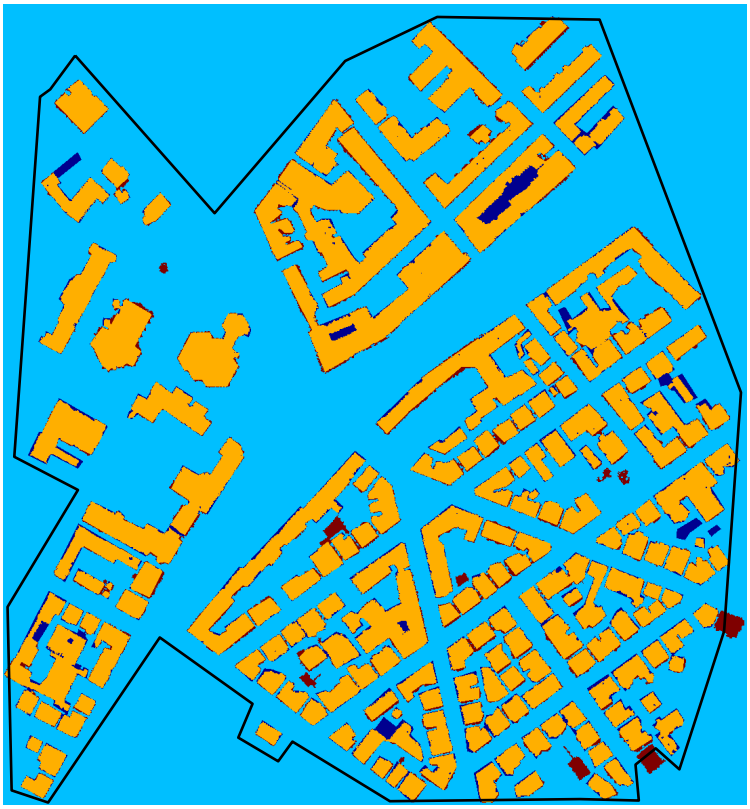


Figure 7.13: Comparison of extracted buildings and corresponding ground truth regions – True positive (Tp - *yellow*), False positive (Fp - *dark blue*), and False negative (Fn - *red*)

$$correctness = 0.92$$

$$quality = 0.86$$

The above mentioned evaluation parameters measured based on difference between the classification regions and the ground truth which are shown as red and dark blue colors in Figure 7.13. Some of the red buildings in particular, those are locating in lower right corner are classified as building but not digitized in ground truth image (cf. Figure 7.12(b)). Two of them are correctly classified (middle and right) but are not digitized because they are not completely inside study area and the left one is part of a large building that is connected to the image border. The small part of this building which is higher than the rest and located inside the image is classified.

Dark blue regions are either not classified or the inner yard of the buildings which are not excluded from the buildings. The not-classified buildings are the low height buildings which are hardly distinguishable even visually as individual buildings.

The next step toward building reconstruction using the approach proposed in this thesis is the extraction of building outlines and their approximation. After classification of the data, the focus is on only the building regions to generate 3D models. The building outlines are extracted from the binary image using morphological erosion. For that, the binary image which contains building pixels (BW) is eroded and the result ($erosionBW$) is subtracted from the (BW):

$$Outlines = BW \text{ - } erosionBW$$

Figure 7.14 shows a result for building outline detection superimposed on the LP image. The scene contains six buildings with six outer boundaries and one inner boundary. For each boundary first the main orientation(s) are extracted using Hough Transform as follows:

- The Hough Transform is computed for the building edges and the main Hough peaks are extracted. The coordinates of the best 10 peaks are extracted if available.
- It is checked if the corresponding angles for the peaks are complementary: If the difference between two peaks is a multiple of 90^o with a possibly small deviation, only one of them is kept.

After determining the main orientations of the buildings, there are two different ways to proceed: If the analysis in Hough space indicates, that there is more than one main orientation, i.e., many main orientations – MMO, the RANSAC-based technique is used (cf. Figure 5.15, right), otherwise the MBR-based method is employed (cf. Figure 5.15, left). Accordingly, outline polygons are extracted and approximated with the RANSAC- or the MBR-based techniques (cf. Sections 5.2.1 and 5.2.2).

Figure 7.15 shows the approximation results for the building boundaries presented in Figure 7.14. They indicate that not only the outer boundaries, but also the inner boundaries can be properly approximated by this method.

After estimating an average height for each building, 3D prismatic models can be formed. If the building contains an inner boundary, its interior will be excluded from the 3D model (cf. Figure 7.16). In Figure 7.17 the 3D prismatic models are combined with the original surface model. Since the average of the height inside each building is used, some (small) parts of the building are not covered by the prismatic model.

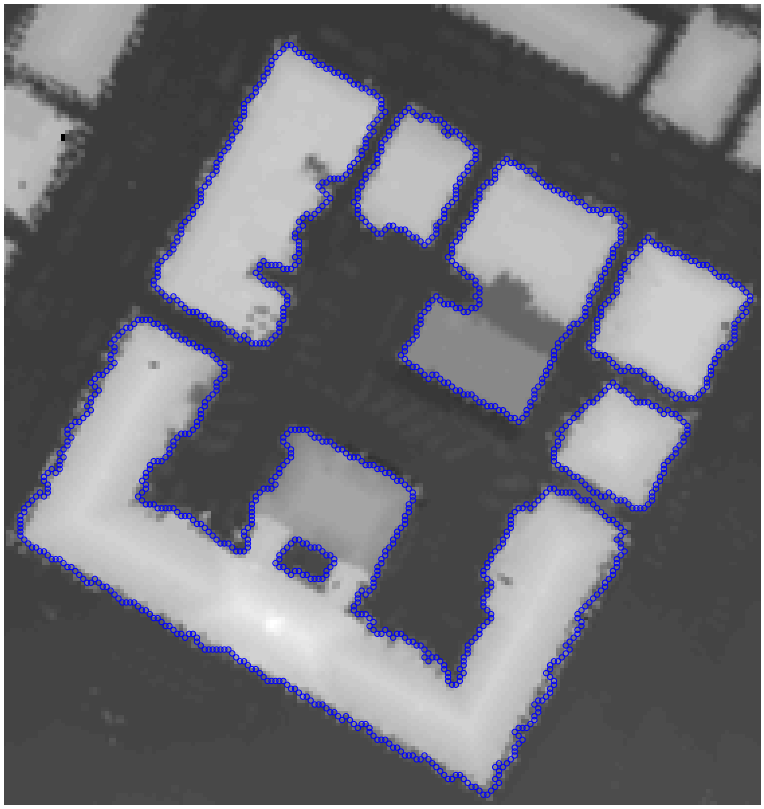


Figure 7.14: Building outlines superimposed on LP

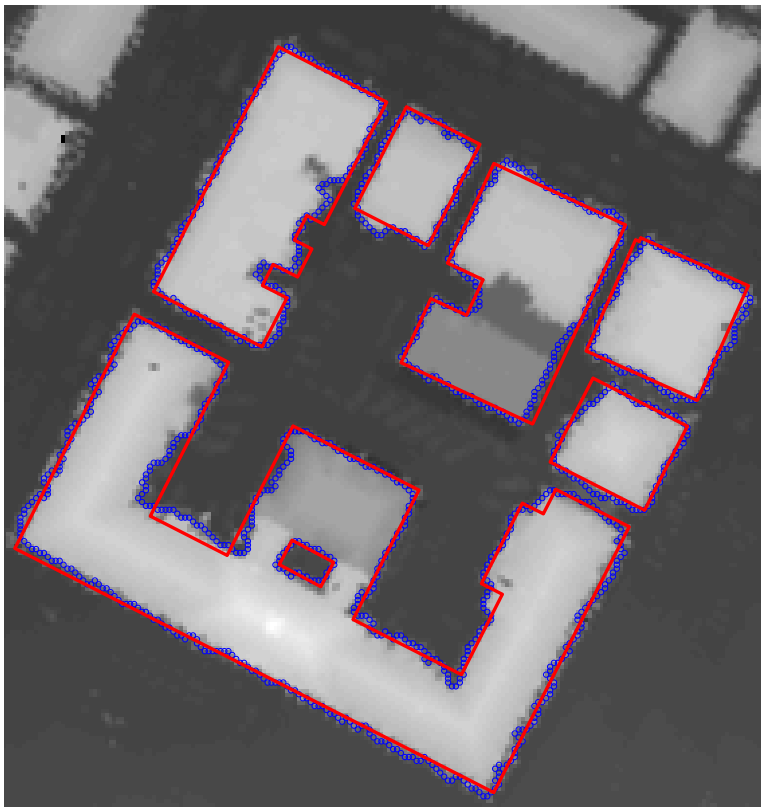


Figure 7.15: Building outline approximation result *(red polygons)* as well as the original extracted building boundaries *(blue polygons)* superimposed on LP image

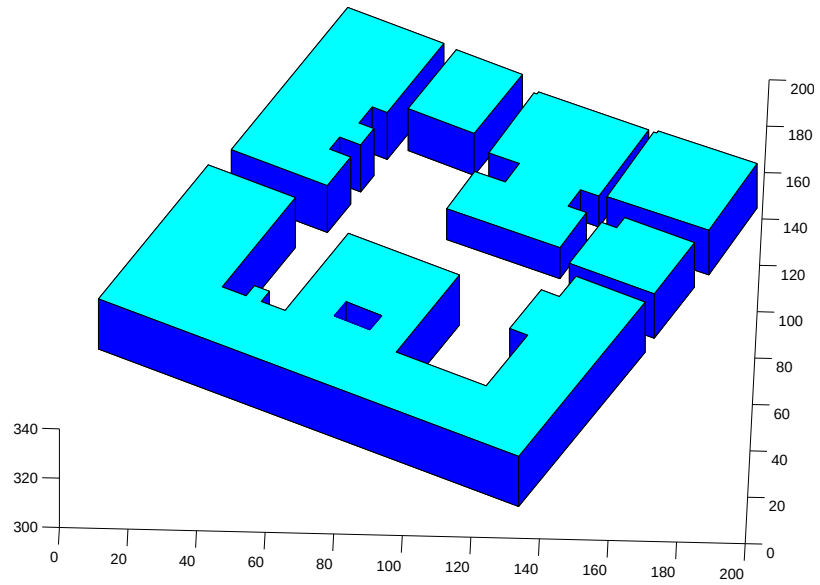


Figure 7.16: 3D prismatic models with average building height

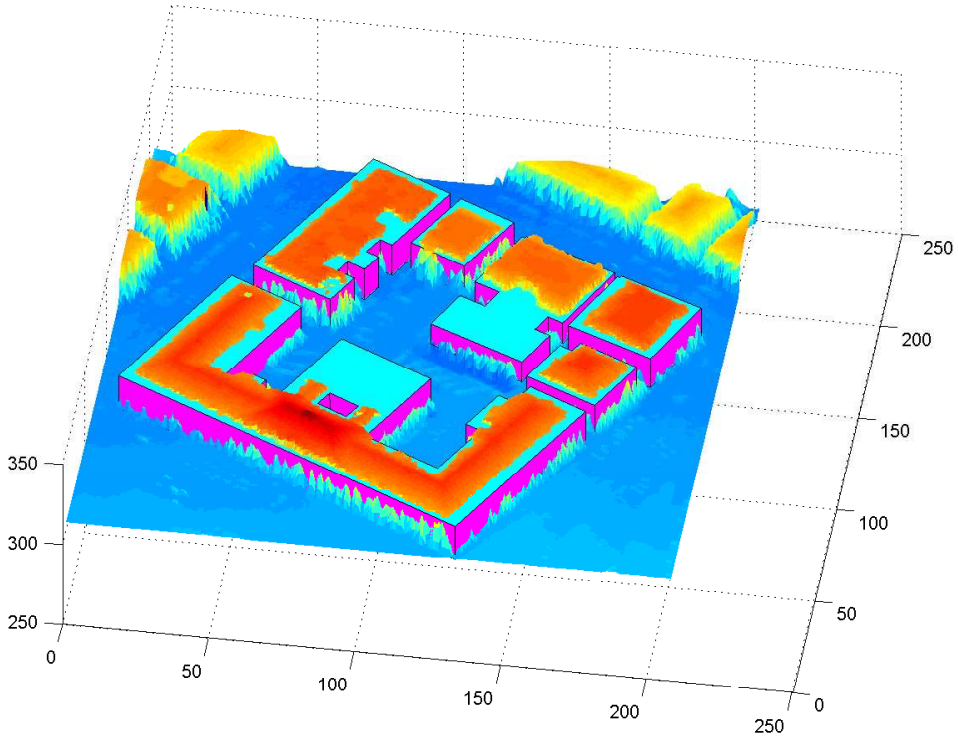


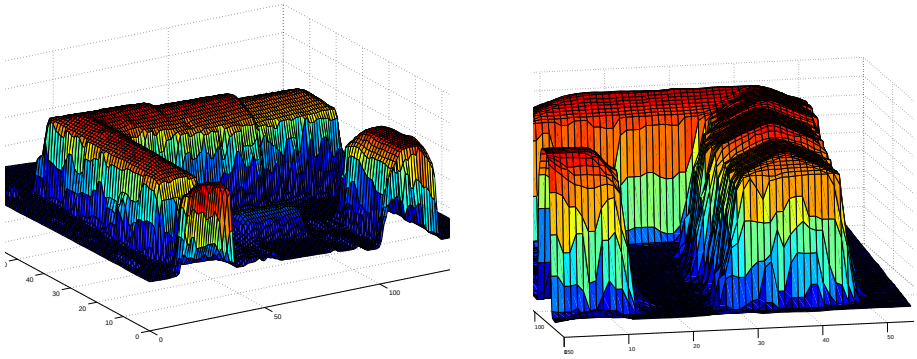
Figure 7.17: 3D prismatic models combined with surface model

7.2.3 3D Building Modeling on LOD2

This section focuses on the evaluation of 3D building models possibly with tilted roofs. They are generated with the novel approach introduced in Section 6.3.4 based on the analysis of the

3D point cloud in projection planes.
Ridge lines are extracted based on geodesic morphological reconstruction. To extract local maxima, the original LP data is chosen as $mask$, and $mask$ minus a small offset h such as $h = 0.5m$ is used as $marker$. The geodesic dilation suppresses the higher parts of the building and other local maxima. After subtracting the reconstructed image from the given image, only the initial local maxima containing the ridge lines remain. By thresholding all pixels corresponding to initial local maxima and the ridge lines are obtained.

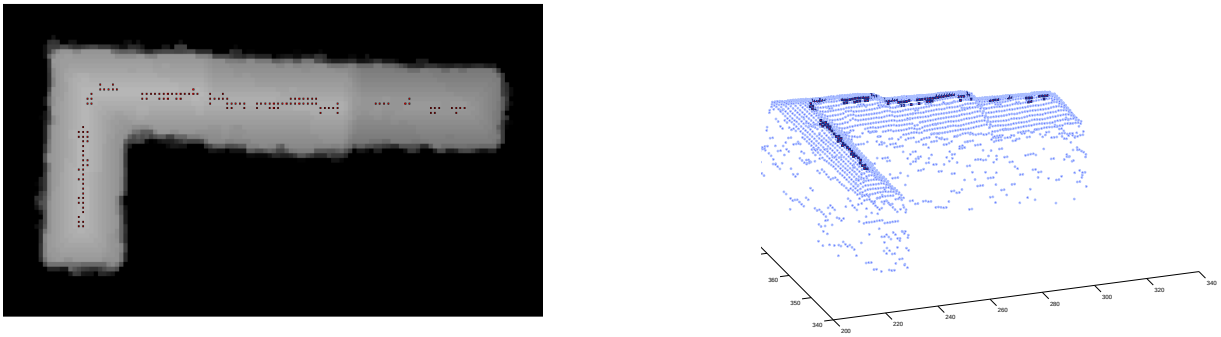
Experiments for an L-shaped building with gable roofs on three different height levels are reported and corresponding models are presented and discussed (cf. Figure 7.18).



(a) (b)

Figure 7.18: L-shaped building in two 3D views — Ridge lines exist on three different height levels.

Points of the initial local maxima extracted by geodesic dilation are presented in Figure 7.19. Their curvatures as well as their 3D surface normal directions are analyzed. Only points with an almost vertical surface normal direction, i.e., for the z component of the normalized surface normal holds $N_z > 0.98$, are retained. Additionally, their corresponding Gaussian and mean curvatures should indicate $flat$ or $ridge - shaped$ surfaces (cf. Section 2.2.3).



(a) 2D view (b) 3D view

Figure 7.19: Extracted initial ridge points indicating different height levels superimposed on LP data

To differentiate between ridge lines and flat surfaces, connected components are labeled. A simple *roundness* criterion is used to separate elongated regions from rectangular or circular regions:

$$roundness = \frac{4 * \pi * area}{perimeter^2}$$

$$roundness \in [0, 1]$$

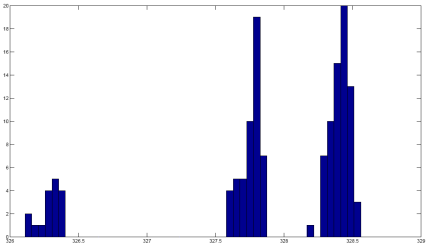
$roundness = 1$ indicates a circular shape and $roundness = 0$ a line.

Regions having a small number of pixels are excluded from this evaluation and they are assumed to be part of ridge lines.

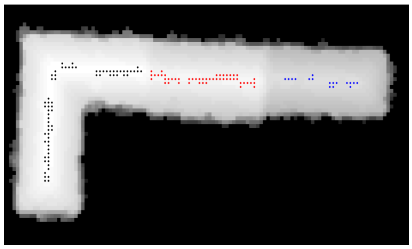
From the ridge points straight line segments having a maximum number of points associated with them are iteratively determined with RANSAC. A threshold of one meter defines the buffer zone around each tested line to find the *Consensus* set of points associated with the line.

It was found empirically, that it is advantageous to classify ridge points based on the histogram of the height values (cf. Figure 7.20(a)) before straight line determination. This is due to, that if the ridge lines are situated on very close height levels, RANSAC has difficulties to detect the individual lines. For the given example two of the ridge lines have a height difference of only $70cm$. This difference is visible in Figure 7.20(a), where the two big peaks on the right side of the histogram correspond to the red and black points in Figure 7.20(b) with a height of $327.8m$ and $328.5m$, respectively.

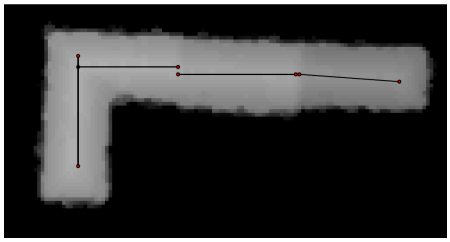
For each ridge line the orientation is calculated and verified by comparing it to the orientation of the boundary lines or the main orientation. If it deviates by less than 5^o from the main orientation or its complement values, the line direction is corrected based on the main orientation (cf. Section 6.3.1).



(a) Histogram of height values



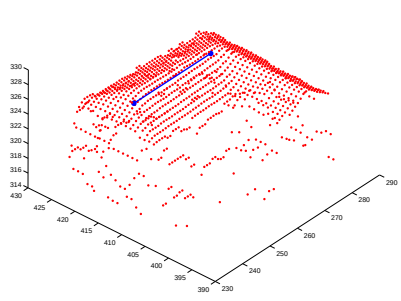
(b) Histogram based classification of the ridge points on three height levels



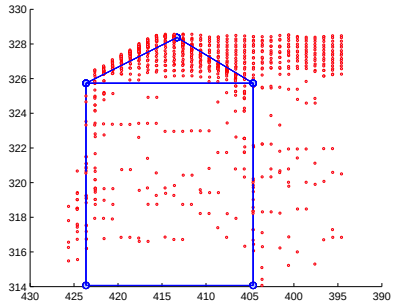
(c) Extracted ridge lines using RANSAC

Figure 7.20: Ridge extraction – Analysis of height values using a histogram (top) and RANSAC-based ridge line extraction (bottom)

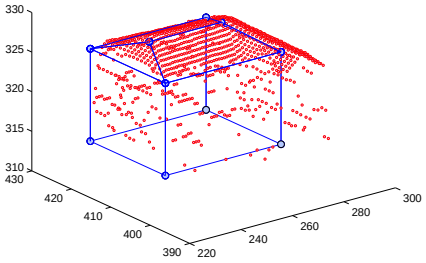
Each ridge line specifies an individual building part and it and its neighboring pixels are analyzed. According to the length of the ridge line, a rectangle with equal distances to both sides of the ridge line is defined in the localization step (cf. Section 6.3). All points included in a cuboid based on the defined rectangle which covers the building part are projected onto a 2D projection plane, being one of the planes of the cuboid. Figure 7.21 represents the 3D model generation for the building part corresponding to the left ridge line in Figure 7.20(c).



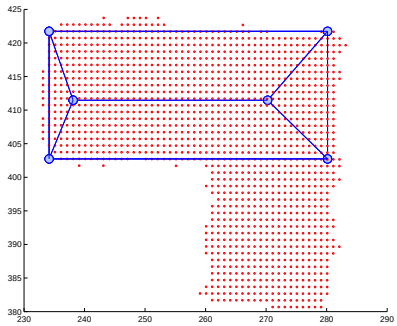
(a) Ridge line and points located in its corresponding cuboid



(b) 2D model of the building part



(c) 3D model – perspective view



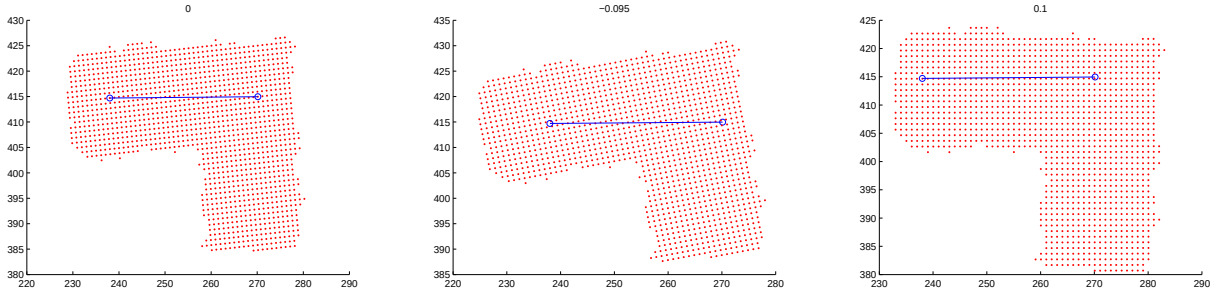
(d) 3D model – top view

Figure 7.21: 3D model generation of the building part for the left ridge line in Figure 7.20(c)

In the 2D projection the location and orientation of the ridge lines are optimized. It is assumed that a correct ridge line produces the maximum density of the point distribution in the 2D view. I.e., the number of cells containing points in the 2D projection (cf. Figure 7.21(b)) is minimum if the direction of the ridge line is exactly parallel to the orientation of the building part. To obtain a high accuracy for the orientation, the localized points are rotated in small steps, e.g., 0.005 rad (ca. 0.3^o), and the number of cells, e.g., with $1m$ size, which contain points in the 2D view is counted. The local orientation with the minimum number of cells is added to the orientation of the ridge line, resulting into an accurate orientation of the building part. Using the correct orientation, the location of the ridge line is refined by moving left and right locating the position with the maximum number of support points.

The localized points are further analyzed to determine the direction of the tilted roofs and the left and right wall. The 3D model of the building part is generated by extending the 2D model to 3D and extracting the front and back walls as well as the direction and location of the tilted roofs (cf. Figures 7.21(c) and 7.21(d)).

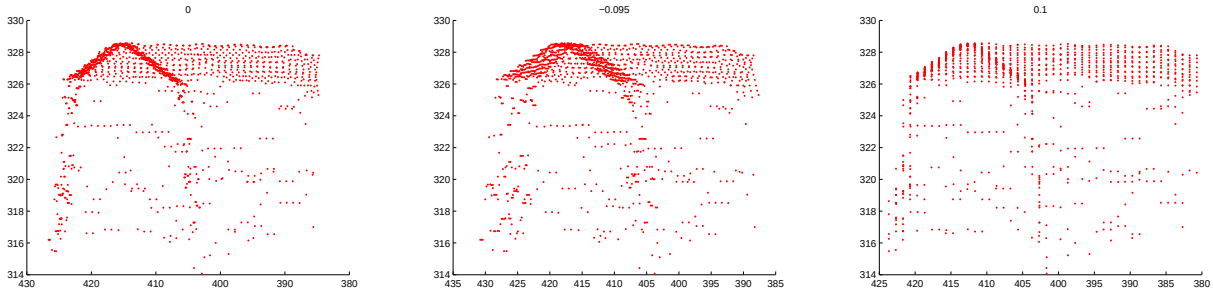
Figure 7.24 shows the final 3D model of the L-shaped building after merging the 3D models of the individual building parts. To aid visual comparison, the merged model is superimposed on



(a) Ridge line computed by RANSAC – orientation correction = 0

(b) Orientation correction = −0.095 *rad*

(c) Orientation correction = 0.1 *rad*

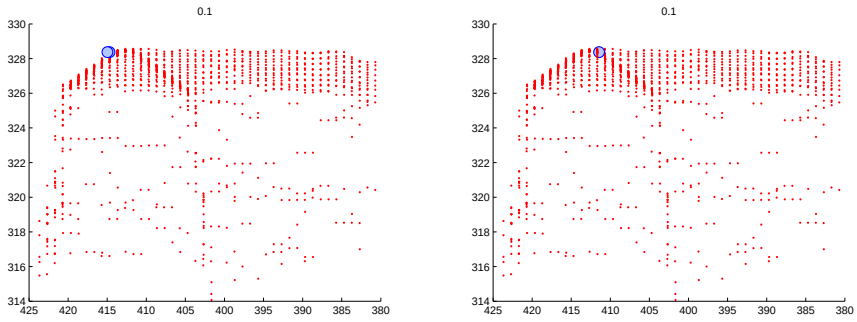


(d) Orientation correction = 0 *rad* – cf. a

(e) Orientation correction = −0.095 *rad* – cf. b

(f) Orientation correction = 0.1 *rad* – correct orientation – cf. c

Figure 7.22: Correction of the main orientation by iterative rotation of the localized points, computing the number of cells containing points in the 2D projection (bottom row)



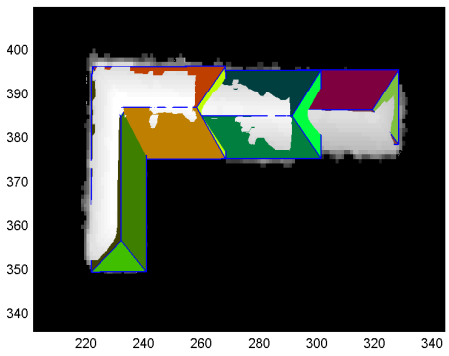
(a) Points *(red)* corresponding to Figure 7.22(f) and position of ridge line *(blue circle)*

(b) Ridge line moved to the correct location

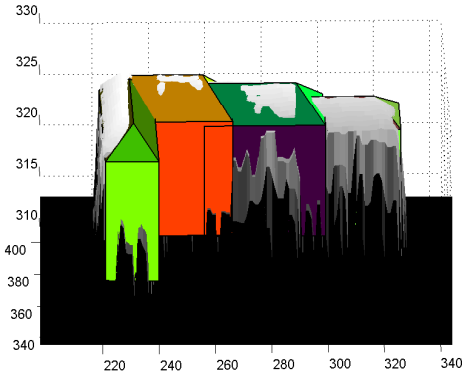
Figure 7.23: Correction of the ridge line location by iteratively shifting the line left and right to find the position with the maximum number of support points

the original laser data. Gray pixels indicate that the model in that region is below the surface.

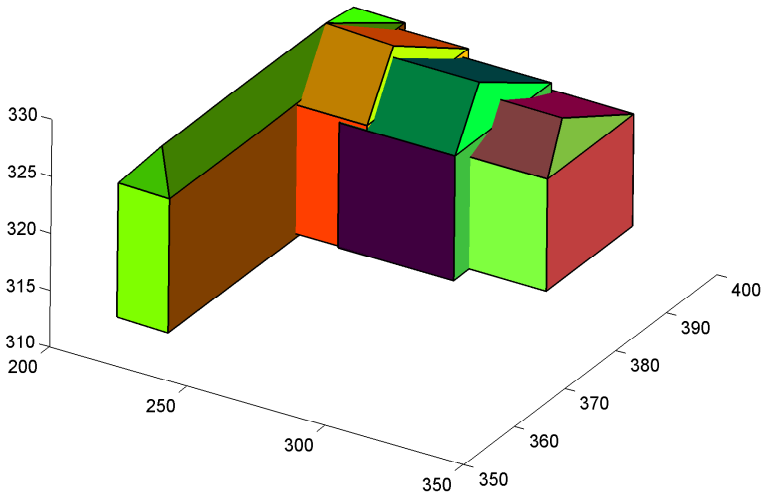
If a building part contains a flat or mansard shaped roof, automatically another technique for modeling is used: As described above, local maxima are classified into ridge points and flat roofs using a roundness criterion. The points at the boundary of the flat roof segment are analyzed if the building part has a flat or a mansard shaped roof. The orientation of the building is defined by the MBR fitted to the building part. If the building has a flat roof, the model is the same as for the LOD1. If the roof is mansard shaped, another rectangle inside and parallel to



(a) Model superimposed on LP data – top view



(b) Model superimposed on LP data – perspective view

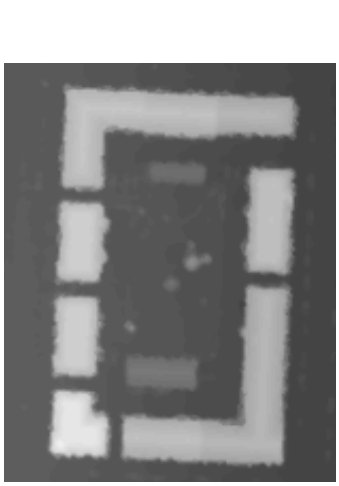


(c) 3D model in perspective view

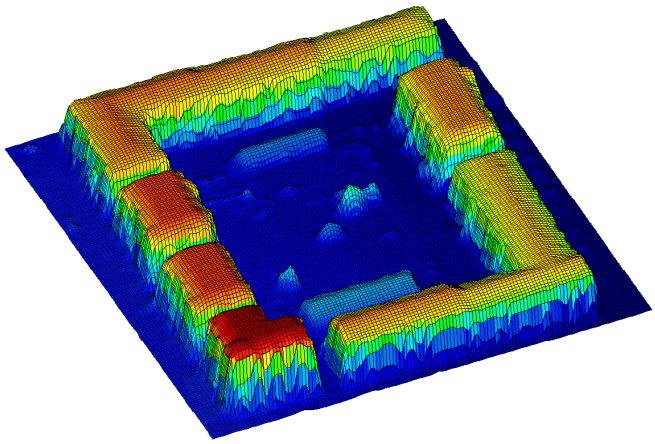
Figure 7.24: Final merged 3D model for L-shaped building

the MBR is fitted to the flat part and the 3D model is formed.

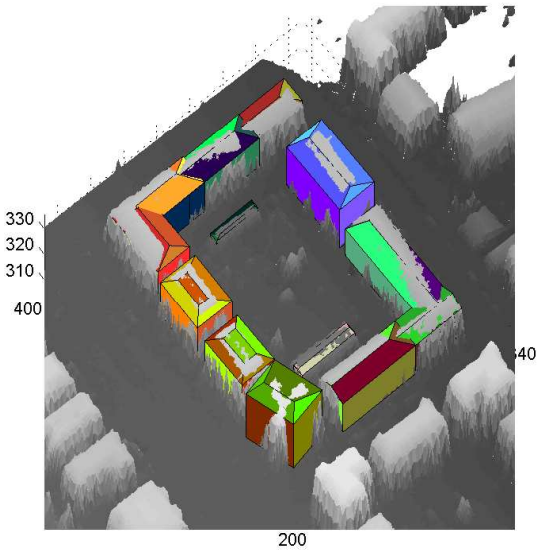
Figure 7.25 represents an area containing eight buildings with different roof shapes: three L-shaped, two mansard shaped, and three hipped roof buildings. Figures 7.25(c) and 7.25(d) show at the lower left corner of the scene a small L-shaped building which is not properly modeled. The ridge points are extracted appropriately using geodesic method. As the two ridge lines contain only few pixels, RANSAC based straight line detection only detected one inclined line instead of two perpendicular lines.



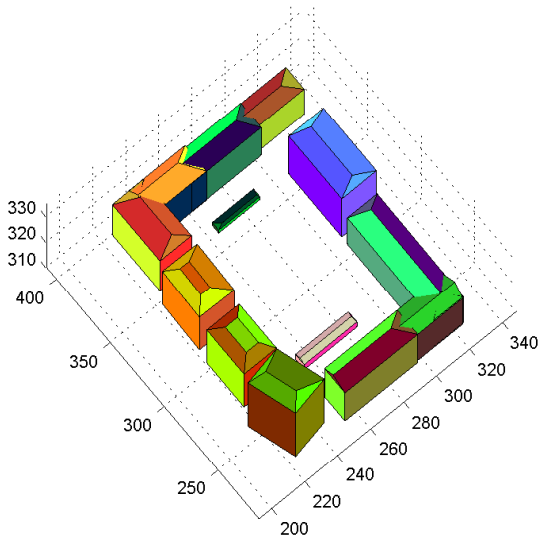
(a) LP image



(b) LP data – perspective view



(c) 3D model superimposed on LP data



(d) 3D model – perspective view

Figure 7.25: 3D building models generated in an area containing eight buildings

7.3 Discussion of Results

In this thesis automatic methods for the production of building models according to the first three LOD of the CityGML standard are explained. Results have been shown in detail in the previous Sections 7.2.1, 7.2.2, and 7.2.3. In this section the results and their strengths and weaknesses are discussed for each LOD:

7.3.1 Generation of LOD0

The assessment of the quality of the procedure for generating the first LOD can be divided into two parts, namely before and after interpolation of gaps, and, therefore, they are discussed separately:

1. Separation of non-ground and ground regions: This part is at the heart of the DTM generation, i.e., LOD0. Iterative geodesic dilation is used to separate above ground objects from the LIDAR data. The provided results show:
   - Figure 7.6(c) indicates that almost all non-ground objects regardless of their size are eliminated.
   - The iterative approach gives good results in hilly as well as smooth residential areas.
   - Profiles 7.8(a), 7.8(b), 7.9(a), and 7.9(b) prove that all objects with large jumps on their boundaries are appropriately eliminated. As shown in Profiles 7.9(a) and 7.9(b) only objects located at the boundary and not-entirely situated inside the image are not deleted.
   - Histogram 7.10 as well as the statistics show that most of the DTM pixels, i.e., 92%, have height differences of less than $1m$ compared to manually derived ground truth.
2. DTM after gap interpolation: The Inverse Distance Weighting (IDW) algorithm is used to fill the gaps. In particular in big gaps located at hills some incorrect values are computed. Even if only few outliers, i.e., non-ground pixels, remain after filtering, interpolation may extend those pixels and severely change the shape of the ground in that area.

7.3.2 Generation of LOD1

Two main parts are included in the generation of the second LOD presented in Section 7.2.2, namely building detection and prismatic model generation:

1. Classification: Rule-based classification is used to classify the 3D objects generated by segmenting the nDSM provided in Section 7.2.1. Feature descriptors such as local range variation (LRV) and variance of the surface normals are used as classification criteria. A new vegetation index (VI) decreases the effects of different object boundary pixels, e.g., of buildings in first- (FP) and last-pulse (LP) LIDAR data. It has been used to extract most of the vegetation regions, i.e., with different height values in the FP and LP data, as well as 3D objects like cranes and power lines. Power lines will only be detected if they are measured by the LIDAR system, which is often not the case in city areas due to their small width. The quality of the classification is assessed as follows:

- Visual interpretation indicates that most of the objects in the study area are correctly classified.
- A statistical evaluation of the classified buildings in comparison to manually acquired ground truth gives a completeness of 93%, a correctness of 92%, and an overall quality of 86% showing that most buildings are appropriately detected.
- Very small as well as elongated buildings with a very small width are the two main building types which cannot be separated from vegetation if corresponding vegetation pixels remain in the LP data. Elongated buildings with a width of a small number of pixels almost show the same values as narrow tree rows in the feature descriptors.

2. Prismatic model generation: A flexible approach for the approximation of rectilinear and non-rectilinear buildings is the basis for the generation of 3D prismatic models. MBR- or RANSAC-based approximation is selected based on the number of main orientations of the building outline. The results shown in Figures 5.16 and 7.15 indicate, that the quality of the approximation is at least acceptable.

   The average height of the pixels inside to a building is used to construct a 3D prismatic model from the approximation result. Figure 7.17 shows that the average height is not a proper representative for the overall building as it is in many cases below or above the height of the building eave.

7.3.3 Generation of LOD2

In this Section the quality of the parametric building models generated by means of projection based 3D modeling (cf. Section 7.2.3) is discussed for each step:

1. The points are projected into the 2D plane based on the direction of the ridge line. Therefore, the quality of the model is directly related to that of the detected ridge line. In LIDAR raster data ridge lines are more pronounced and can thus be more accurately and reliably determined than building edges. Since there exist almost no discontinuities for the pixels around ridge lines, the ridge lines and their neighboring pixels are not negatively affected by interpolation. Additionally, contrary to the building edges where the locations of the pixels are slightly different in FP and LP data, the pixels around the ridge line have the same height values in FP and LP data. The initial ridge points are detected using geodesic dilation and are evaluated based on surface normal direction and curvature values. As shown in Figures 6.5, 6.6, and 7.20(b), the quality of the points classified as ridge is appropriate and they are located at the correct positions. Yet, Figures 6.5 and 6.6 also gives some points detected by geodesic dilation as local maxima which belong to small structures above the roof.

2. RANSAC-based ridge line detection is a sensitive process to extract the straight line representing the ridge line of the building in the correct orientation. The parameters of RANSAC, particularly the threshold for the distance to accept points as inliers, have to be carefully tuned. On one hand, it should be large enough to extract not too many straight lines, but on the other hand, it should be small enough to separate ridge lines situated at a small height difference, as can be seen in Figure 7.20(c). Thus, to avoid having to deal with ridge lines located very close to each other, they are classified in advance using histogram analysis (cf. Figure 7.20(a)). This step solved the problem of ridge lines with small differences in their heights. Yet, as can be seen for the small L-shaped building in

the lower left corner in Figure 7.25, if the ridge lines are small, close to each other, and have the same height, RANSAC cannot separate the lines. In this case, only one diagonal line instead of two perpendicular lines is extracted.

3. The orientation of the building part could be optimized by iteratively rotating the points thought to belong to the building and counting the number of cells containing points in the 2D projection (cf. Figure 7.22).

4. The locations of the ridge lines could also be optimized by moving them left and right and counting the number of supporting points (cf. Figure 7.23).

5. A mansard shaped building has been modeled after classifying the roof top pixels as partially flat and partially inclined (cf. Section 7.2.3). Two parallel MBR define the roof faces.

Chapter 8

Conclusions and Recommendations

8.1 Summary and Conclusions

In this thesis a comprehensive workflow to generate 3D building models from LIDAR point clouds is proposed. The point clouds acquired by airborne laser scanning systems are analyzed to produce 3D models according to the definitions of the CityGML standard. Geodesic morphological processing is at the core of the algorithms, e.g., for filtering of non-ground regions and ridge point extraction.

The automated 3D building reconstruction from irregular LIDAR point clouds begins with preprocessing consisting of the elimination of outliers, data rasterization, and the removal of gaps. The resulting raster LIDAR data is analyzed to separate the non-ground from the ground regions. Geodesic dilation is employed to hierarchically filter the non-ground regions. By interpolating of the gaps, digital terrain models (DTM) are generated. The 3D representation of the ground regions as DTM forms the first level of detail, i.e., LOD0.

The normalized digital surface model (nDSM) provided by the comparison between the DTM and the LIDAR DSM is used as basis for the segmentation and classification of 3D objects. A single threshold and connected component labeling produce initial 3D regions which are evaluated based on feature descriptors to extract and recognize building and vegetation areas. The procedure is employed on last pulse (LP) height data and, therefore, the vegetation areas only comprise dense vegetation regions still available in LP. Because most of the vegetation regions are visible only in first pulse (FP) data, a vegetation index (VI) is computed by comparing FP and LP and employing morphological dilation. This feature descriptor highlights most vegetation regions, but also cranes, thick power lines, and all other objects which contain small holes or parts thinner than the diameter of the laser footprint.

The outlines of building regions are approximated (if possible) by rectilinear polygons which are the basis for the second LOD or LOD1 by an MBR- and a RANSAC-based method. The former is model-driven, hierarchically fitting minimum bounding rectangles (MBR) to the building outline. In contrary, the RANSAC-based method is data-driven, attempting to extract straight lines from the building outline and merging them to form a closed polygon. The approximation automatically decides which of the methods should be used based on the main orientation of the building extracted by Hough Transform. If a building has one main orientation (OMO), the MBR-based method, and if there exist many main orientations (MMO), the RANSAC-based technique is employed. For the approximated outlines an average height value is calculated based on the internal points of each building leading to a prismatic building model forming the LOD1.

To form the third LOD or LOD2, a projection based approach is proposed reconstructing a building model with roof structures. The reconstruction to extract the building ridge lines employs a small value to generate the *marker* image used to determine the local maxima of the image. Thresholding leads to hypotheses for ridge points evaluated by values for the surface normal as well as mean and Gaussian curvature. Connected component analysis provides initial regions which are classified into flat roofs and ridge lines based on roundness.

The 3D points of the ridge lines are analyzed by RANSAC to extract straight lines. For each ridge line one parametric model is generated. The models of the building parts are merged finally.

The different steps of the proposed methodology for the automatic 3D reconstruction of buildings from LIDAR point clouds can be characterized as follows:

1. Filtering of the laser data using geodesic reconstruction:

   (a) The result is independent of the size of the objects. Thus spacious as well as small buildings can be efficiently filtered using this approach.

   (b) Contrary to standard morphological processing, the definition of proper structuring elements is not necessary. In geodesic dilation the marker image is dilated by an elementary isotropic structuring element.

   (c) As geodesic reconstruction does not affect ground pixels, the normalized DSM can simply be segmented using a threshold of *zero*.

   (d) Filtering based on geodesic dilation is relatively fast. In many cases even in hilly regions only a single marker image is needed representing the minimum height value of the *mask* image except for pixels at the boundary where $marker = mask$ is used.

   (e) Outliers, either below or above the terrain, do not have a negative effect on the filtering, as for basic gray-scale dilation or erosion.

   (f) A drawback is, that objects, which are not entirely located inside the image, cannot be filtered.

2. Classification of objects and approximation of building outlines:

   (a) The vegetation Index (VI) proposed in this thesis can largely eliminate negative effects at the building edges and it can thus improve the quality of the extracted regions.

   (b) Vegetation regions can be separated from high linear objects such as cranes and power lines by evaluating the height variations of the original laser points of each region.

   (c) A disadvantage of the vegetation extraction method is, that vegetation regions which have almost the same FP and LP value cannot always be extracted. It can also happen that they have the same height variation as buildings.

   (d) The method to be used for approximation can be automatically determined by analyzing the main orientations of the building.

   (e) The MBR-based method is inherently model-driven and hierarchical. It can thus be adapted easily to different generalization levels. The approximated polygons form due to their orthogonality an excellent basis for 3D building modeling.

(f) A further advantage of the MBR-based method is the good preservation of symmetries.

(g) The RANSAC-based method is flexible and can be used for most of the building outlines which are not rectangular.

3. Projection-based 3D reconstruction of buildings:

(a) Parametric building reconstruction is model-driven and begins with extracting ridge lines used to reconstruct gable roofs.

(b) For big buildings even dormers with ridge lines can be modeled. It is also possible for smaller buildings in point clouds with a higher density.

(c) By extracting the ridge lines and projecting the corresponding points for each building part into the horizontal 2D plane, the model can be appropriately fitted.

(d) Strengths of the projection-based approach are its robustness and, that it is quite fast, as projection into 2D space reduces the algorithmic complexity significantly.

8.2 Recommendations

1. In this study a LIDAR data set with a density of about 4.8 points per m^2 has been used as basis for the analysis of the proposed approaches for the classification of surface pixels and the 3D modeling of buildings. Even though this point density is sufficient for separating non-ground and ground points, a higher density might improve the recognition and 3D modeling of buildings. In Section 7.3.2 it was shown that small and elongated buildings with small widths cannot be appropriately discriminated from vegetation. For such buildings the feature descriptors namely the variances of the surface normals of the pixels inside a region show the same values as for single trees or rows of trees. A higher point density might improve the quality of the feature descriptors for the evaluation of the internal pixels.

2. Although a higher point density might improve the classification, geometrical information alone might in many cases still not be sufficient for a highly reliable classification. Thus, other data sources such as color aerial photos or multi-spectral data should be considered.

3. For the prismatic model for buildings the average height of the internal pixels might not be a proper representative particularly for building blocks with parts on different height levels. For this case it should be better to first (iteratively) segment using different offsets h all building parts on the same height level and then find walls between them.

4. The generation of parametric building models based on the extraction of ridge lines or roof top pixels is novel and has many advantages compared to 3D reconstruction methods detecting and merging planar faces. The first and arguably biggest benefit is that the building can easily be divided into building parts based on the assumption that "every ridge line represents one building part". The second advantage is the reduction of the complexity by projection from 3D into 2D space using the location and orientation of the ridge line, i.e., lines can be analyzed instead of planes.
Though only building parts with tilted and mansard shaped roofs have been extracted in this thesis, the approach can be extended to other roof types. The approach is simple, straightforward, and fast. It should be possible to implement it for most of the buildings

in suburban areas with not too complex structures. Yet, the result depends on the quality of the extracted ridge lines and flat roof pixels. Only robust techniques which extract ridges with high quality can guarantee the quality of the building parts. In this thesis the ridges are in most cases appropriately extracted, but as shown, e.g., in Figure 7.25, if the ridge is small and consists of more than one line, the approach can fail to detect the correct lines.

5. The objective of this thesis was to present the potential of ridge and projection based analysis of LIDAR data for 3D building model generation. Though the experiments have basically verified this objective, refinements and adjustments are required to make the approach more flexible and powerful enough to deal with larger amounts of more general data on a quality level sufficient for practical applications.

Bibliography

Ackerman, F. and Krzystek, P. (1991): MATCH-T: Automatic mensuration of digital elevation models, *3rd Technical Seminar of the Sociedad Espanola de Cartografia, Fotogrametria y Teledeccion*, Barcelona.

Alharthy, A. and Bethel, J. (2002): Heuristic filtering and 3D feature extraction fromLIDAR data, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 34 (3A), 29 – 34.

Ameri, B. (2000): *Automatic Recognition and 3D Reconstruction of Buildings from Digital Imagery*, PhD thesis: Deutsche Geodätische Kommission, C 526.

Arefi, H. and Hahn, M. (2005): A morphological reconstruction algorithm for separating off-terrain points from terrain points in laser scanning data, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 36 (3/W19).

Arefi, H., Engels, J., Hahn, M. and Mayer, H. (2007a): Approximation of building boundaries, *Urban Data Management Systems (UDMS) Workshop, Stuttgart*, 25 – 33.

Arefi, H., Engels, J., Hahn, M. and Mayer, H. (2007b): Automatic DTM generation from laser-scanning data in residential hilly area, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 36 (4/W45).

ASPRS (2004): ASPRS, Guidelines – vertical accuracy reporting for LiDAR data, Version 1, *Technical report*, American Society for Photogrammetry and Remote Sensing.

Axelsson, P. (1999): Processing of laser scanner data - algorithms and applications, *ISPRS Journal of Photogrammetry and Remote Sensing* **54**: 138–147.

Baillard, C., Schmid, C., Zisserman, A. and Fitzgibbon, A. (1999): Automatic line matching and 3D reconstruction of buildings from multiple views, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 32 (3-2W5), 69 – 80.

Baltsavias, E. P. (1999): A comparison between photogrammetry and laser scanning, *ISPRS Journal of Photogrammetry and Remote Sensing* **54**: 83–94.

Baltsavias, E. P., Grün, A. and Van Gool, L. (2001): *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*, Taylor & Francis.

Bartels, M. and Wei, H. (2006): Segmentation of LIDAR data using measures of distribution, *International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, Volume 36 (7), 426 – 431.

Besl, P. J. (1988): *Surfaces in Early Range Image Understanding*, PhD thesis, University of Michigan.

Besl, P. and McKay, N. (1992): A method for registration of 3-d shapes, *Pattern Analysis and Machine Intelligence* **14**(2): 239 – 256.

Brenner, C. (2000a): *Dreidimensionale Gebäuderekonstruktion aus digitalen Oberflächenmodellen und Grundrissen*, PhD thesis: Deutsche Geodätische Kommission, C 530.

Brenner, C. (2000b): Towards fully automatic generation of city models, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 33 (B3), 85 – 92.

Brunn, A. and Weidner, U. (1997): Extracting buildings from digital surface models, *IAPRS*, 3–4.

Burns, J., Hanson, A. and Riseman, E. (1986): Extracting straight lines, *Pattern Analysis and Machine Intelligence* **8**(4): 425 – 455.

Clode, S., Kootsookos, P. and Rottensteiner, F. b. (2004): Accurate building outlines from ALS data, *12th Australasian Remote Sensing and Photogrammetry Conference*, Perth, Western Australia.

Douglas, D. and Peucker, T. (1973): Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *The Canadian Cartographer* **10**(2): 112–122.

Elaksher, A., Bethel, J. and Mikhail, E. M. (2002): Building extraction using multiple images, *ACSM-ASPRS 2002 Annual Conference*.

Fischler, M. and Bolles, R. (1981): RAndom Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM* **24**(6): 381–395.

Flynn, P. and Jain, A. (1988): Surface classification: Hypothesis testing and parameter estimation, *Computer Vision and Pattern Recognition*, 261 – 267.

Förstner, W. (1999): 3D City Models: Automatic and Semiautomatic Acquisition Methods, *Photogrammetric Week, University of Stuttgart*, 291–303.

Forsyth, D. A. and Ponce, J. (2003): *Computer Vision: A Modern Approach*, Prentice Hall.

Fuchs, C. (1998): *Extraktion polymorpher Bildstrukturen und ihre topologische und geometrische Gruppierung*, PhD thesis: Deutsche Geodätische Kommission, C 502.

Geibel, R. and Stilla, U. (2000): Segmentation of laser-altimeter data for building reconstruction: Comparison of different procedures, *International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, Volume 33 (B3), 326 – 334.

Gonzalez, R. C., Woods, R. E. and Eddins, S. L. (2004): *Digital Image Processing using Matlab*, Prentice Hall, Upper Saddle River, NJ.

Gonzalez and Woods (2008): *Digital Image Processing*, Prentice Hall, Upper Saddle River, NJ.

Gorte, B. (2002): Segmentation of TIN-structured surface models, *International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, Volume 34 (4).

Grubbs, F. E. (1969): Procedures for detecting outlying observations in samples, *Technometrics* **11**(1): 1–21.

Gülch, E., Müller, H. and Hahn, M. (2004): Semi-automatic object extraction – lessons learned, *International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, Volume 35 (B3), 488 – 493.

Haala, N. and Anders, K.-H. (1997): Acquisition of 3D urban models by analysis of aerial images, digital surface models and existing 2D building information, *SPIE Conference on Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III* **3072**: 212 – 222.

Haala, N. and Brenner, C. (1999): Extraction of buildings and trees in urban environments, *ISPRS Journal of Photogrammetry & Remote Sensing* **54**: 130 – 137.

Hartley, R. and Zisserman, A. (2004): *Multiple view geometry in computer vision*, Cambridge University Press, Cambridge, UK.

Hebel, M. and Stilla, U. (2007): Automatic Registration of Laser Point Clouds of Urban Areas, *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, Volume 36(3 - W49A), 13 – 18.

Hebel, M. and Stilla, U. (2008): Pre-classification of Points and Segmentation of Urban Objects by Scan Line Analysis of Airborne LIDAR data, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 37 (3 - B3a), 105 – 110.

Heipke, C., Mayer, H., Wiedemann, C. and Jamet, O. (1997): Evaluation of automatic road extraction, *International Archives of Photogrammetry and Remote Sensing*, Volume 32 (3-2W3), 47 – 56.

Hoffman, R. and Jain, A. (1987): Segmentation and classification of range images, *Pattern Analysis and Machine Intelligence* **9**(5): 608 – 620.

Hoover, A., Jean Baptiste, G., Jiang, X., Flynn, P., Bunke, H., Goldgof, D., Bowyer, K., Eggert, D., Fitzgibbon, A. and Fisher, R. (1996): An experimental comparison of range image segmentation algorithms, *Pattern Analysis and Machine Intelligence* **18**(7): 673 – 689.

Hough, P. (1962): Methods and means for recognizing complex patterns, *U.S. Patent 3,069,654*.

Hu, J., You, S., Neumann, U. and Park, K. K. (2004): Building modeling from LiDAR and aerial imagery, *Proceedings of the ASPRS*.

INTERMAP Technologies (2008): Definition of Digital Terrain Models (DTMs), www.intermap.com/right.php/pid/3/sid/15/tid/15: Online; Last access 10 April 2008.

Jähne, B., Haussecker, B. and Geissler, P. (1999a): *Handbook of Computer Vision and Application, Volume II, Signal Processing and Pattern Recognition*, Academic Press.

Jähne, B., Haussecker, B. and Geissler, P. (1999b): *Handbook of Computer Vision and Application, Volume III, Systems and Applications*, Academic Press.

Jain, A. and Dubes, R. C. (1988): *Algorithms for Clustering Data*, Prentice Hall, Englewood Cliffs, NJ.

Jiang, X. and Bunke, H. (1994): Fast segmentation of range images by scan line grouping, *Machine Vision and Applications* **7**(2): 115 – 122.

Jutzi, B. and Stilla, U. (2006): Range determination with waveform recording laser systems using a Wiener Filter, *ISPRS Journal of Photogrammetry and Remote Sensing* **61 (2)**: 95 – 107.

Kilian, J., Haala, N. and Englich, M. (1996): Capture and evaluation of airborne laser scanner data, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 31 (B3), 383–388.

Kirchhof, M., Jutzi, B. and Stilla, U. (2007): Iterative processing of laser scanning data by full waveform analysis, *ISPRS Journal of Photogrammetry and Remote Sensing* **63**: 99 – 114.

Klein, L. A. (1999): *Sensor and Data Fusion Concepts and Applications (Tutorial Texts in Optical Engineering)*, SPIE-International Society for Optical Engine.

Kolbe, T. H., Gröger, G. and Plümer, L. (2005): CityGML - Interoperable Access to 3D City Models, *International Symposium on Geo-information for Disaster Management*, Springer Verlag.

Kraus, K. and Pfeifer, N. (1998): Determination of terrain models in wooded areas with airborne laser scanner data, *ISPRS Journal of Photogrammetry and Remote Sensing* **53**: 193–203.

Kraus, T., Lehner, M. and Reinartz, P. (2007): Modeling of urban areas from high resolution stereo satellite images, *Proceedinsg of the ISPRS Hannover Workshop 2007 High-Resolution Earth Imaging for Geospatial Information.*

Lowe, D. (2004): Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* **60**(2): 91 – 110.

Ma, R. (2004): *Building Model Reconstruction from LIDAR Data and Aerial Photographs*, PhD thesis, The Ohio State University.

Maas, H.-G. (1999a): Fast determination of parametric house models from dense airborne laserscanner data, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 32 (2W1), 1 – 6.

Maas, H.-G. (1999b): The potential of height texture measures for the segmentation of airborne laserscanner data, *Fourth International Airborne Remote Sensing Conference*, Ottawa, 154–161.

Maas, H.-G. (2003): Planimetric and height accuracy of airborne laserscanner data: User requirements and system performance, *Proceedings 49. Photogrammetric Week, Wichmann Verlag*, 117–125.

Martinez, W. L. (2004): *Exploratory Data Analysis with MATLAB (Computer Science and Data Analysis)*, Chapman & Hall/CRC.

Mathworks (2007): MATLAB, the Language of Technical Computing, www.MathWorks.com/products/matlab: Last access February 2007.

Mayer, H. (1999): Automatic object extraction from aerial imagery: A survey focusing on buildings, *Computer Vision and Image Understanding* **74**(2): 138 – 149.

Mayer, H. (2008): Object extraction in photogrammetric computer vision, *ISPRS Journal of Photogrammetry and Remote Sensing* **63**(2): 213 – 222.

McGlone, C., Mikhail, E. and Bethel, J. (Herausgeber) (2004): *Manual of Photogrammetry*, American Society for Photogrammetry and Remote Sensing.

McIntosh, K. and Krupnik, A. (2002): Integration of laser-derived DSMs and matched image edges for generating an accurate surface model, *ISPRS Journal of Photogrammetry and Remote Sensing* **56**(3): 167 – 176.

Mikhail, E. M., Bethel, J. S. and McGlone, J. C. (2001): *Introduction to modern photogrammetry*, John Wiley and Sons, New York.

Müller, H. (1998): Experiences with semiautomatic building extraction, Third Course in Digital Photogrammetry, Bonn, Germany. Institute for Photogrammetry at Bonn University and Landesvermessungsamt Nordrhein-Westfalen.

Nevatia, R., Lin, C. and Huertas, A. (1997): A system for building detection from aerial images, *Monocular Building Detection, Ascona Workshop II*, Ascona Switzerland.

Nister, D. (2004): An efficient solution to the five-point relative pose problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(6): 756 – 770.

NIST/SEMATECH (2007): e-handbook of statistical methods, www.itl.nist.gov/div898/handbook: Last access December 2007.

Optech, I. (2007): www.optech.ca, Visited February 2007: Last access February 2007.

Ramer, U. (1972): An iterative procedure for the polygonal approximation of plane curves, *Computer Graphics and Image Processing* **1**: 244 – 256.

Rieger, W., Kerschner, M., Reitert, T. and Rottensteiner, F. (1999): Roads and buildings from laser scanner data within a forest enterprise, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 32 (3-W14), 185 –191.

Rottensteiner, F. (2006): Consistent estimation of building parameters considering geometric regularities by soft constraints, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 36 (3), 13 – 18.

Rottensteiner, F. and Briese, C. (2002): A new method for building extraction in urban areas from high-resolution LIDAR data, *International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, Volume 34 (3A), 295 – 301.

ROTTENSTEINER, F. and JANSA, J. (2002): Automatic extraction of buildings from LIDAR data and aerial images, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 34, 569–574.

ROTTENSTEINER, F., TRINDER, J., S., C. and KUBIK, K. (2004a): Fusing airborne laser scanner data and aerial imagery for the automatic extraction of buildings in densely built-up areas, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 35 (B3), 512 – 517.

ROTTENSTEINER, F., TRINDER, J., S., C., KUBIK, K. and LOVELL, B. (2004b): Building Detection by Dempster-Shafer Fusion of LIDAR Data and Multispectral Aerial Imagery, *17th International Conference on Pattern Recognition (ICPR)*, Volume 2, 339 – 342.

SAMADZADEGAN, F., AZIZI, A., HAHN, M. and LUCAS, C. (2005): Automatic 3D object recognition and reconstruction based on neuro-fuzzy modelling, *ISPRS Journal of Photogrammetry and Remote Sensing* **59**(5): 255 – 277.

SCHENK, T. (2004): Airborne laser scanning, lecture notes, Stuttgart University of Applied sciences.

SCHENK, T. (2007): Airborne laser scanning systems, lecture notes, Ohio State University.

SCHENK, T. and CSATHO, B. (2002): Fusion of LIDAR Data and Aerial Imagery for a More Complete Surface Description, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 34 (3A), 310 – 317.

SCHMID, C. and ZISSERMAN, A. (1997): Automatic line matching across views, *Computer Vision and Pattern Rwcognition*, 666 – 671.

SCOTT, G. and LONGUET-HIGGINS, H. (1991): An algorithm for associating the features of two patterns, *Royal Society London* **B244**: 21 – 26.

SHAMOS, M. (1978): *Computational Geometry*, PhD thesis, Yale University, Ann Arbor, MI.

SHAPIRO, L. G. and STOCKMAN, G. C. (2001): *Computer Vision*, Prentice Hall, Englewood-Cliffs, New Jersey.

SITHOLE, G. (2001): Filtering of laser altimetry data using a slope adaptive filter, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 34 (3-W4), 203–210.

SITHOLE, G. and VOSSELMAN, G. (2003): Comparison of fitering algorithms, *International Archives of Photogrammetry and Remote Sensing*, Volume 34 (3-W13), ISPRS, 71–78.

STEINLE, E. and VÖGTLE, T. (2000): Effects of different laser scanning modes on the results of building recognition and reconstruction, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 33 (P3), 858 – 865.

STEINLE, E. and VÖGTLE, T. (2001): Automated Extraction and Reconstruction of Buildings in Laser Scanning Data for Disaster Management, *Automatic Extraction of Man-Made Objects from Aerial and Space Images III*, 309 – 318.

STILLA, U. (2007): High resolution radar imaging of urban areas, *Photogrammetric Week, 07. Wichmann: Heidelberg*, 149 – 158.

Stilla, U. and Jurkiewicz, K. (1999): Reconstruction of building models from maps and laser altimeter data, *Integrated Spatial Databases: Digital Images and GIS*, Springer-Verlag, 34–46.

Stilla, U. and Jutzi, B. (2008): Waveform analysis for small- footprint pulsed laser systems, *Topographic Laser Ranging and Scanning: Principles and Processing.*, 215 – 234.

Stilla, U., Söergel, U. and Thönnessen, U. (2003): Potential and limits of insar data for building reconstruction in built up-areas, *ISPRS Journal of photogrammetry and remote sensing* **58(1-2)**: 113 – 123.

Stilla, U., Söergel, U., Thönnessen, U. and Michaelsen, E. (2001): Segmentation of lidar and insar elevation data for building reconstruction, *In: Gruen A, Baltsavias EP, vanGool L (eds) Automatic extraction of man-made objects from aerial and space images (III), Lisse: Balkema*, 297 – 307.

Stilla, U., Yao, W. and Jutzi, B. (2007): Detection of weak laser pulses by full waveform stacking, *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, Volume 36(3/W49A), 25 – 30.

Tarsha Kurdi, F., Landes, T., Grussenmeyer, P. and Koehl, M. (2007): Model-driven and data-driven approaches using LIDAR data: Analysis and comparison, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 36 (3-W49A), 87 – 92.

Terrasolid (2007): Processing laser and images, www.terrasolid.fi: Terrascan software for processing of laser data.

TopScan (2007): TopScan GmbH, www.topscan.de: Last access February 2007.

Vincent, L. (1993): Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms, *IEEE Transactions on Image Processing* **2**: 176–201.

Vincent, L. (1997): *Current Trends in Stochastic Geometry and its Applications*, Chapman & Hall, London, chapter Current Topics in Applied Morphological Image Analysis.

Vögtle, T. and Steinle, E. (2003): On the quality of object classification and automated building modelling based on laser scanning data, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 34 (B-W13), 149 –155.

von Hansen, W. and Voegtle, T. (1999): Extraktion der Geländeoberfläche aus flugzeuggetragenen Laserscanner-Aufnahmen, *Photogrammetrie - Fernerkundung - Geoinformation (PFG)* **4**: 229 – 236.

Vosselman, G. (2000): Slope based filtering of laser altimetry data, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 33 (B3), 935–942.

Vosselman, G. and Dijkman, S. (2001): 3D building model reconstruction from point clouds and ground plans, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 34, 37–44.

Vosselman, G., Gorte, B. and Sithole, G. (2004): Change detection for updating medium scale maps using laser altimetry, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 34 (B3), 207 –212.

Wack, R. and Wimmer, A. (2002): Digital terrain models from airborne laser scanner data - a grid based approach, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 34 (B3), 293 – 296.

Weidner, U. and Förstner, W. (1995): Towards automatic building extraction from high resolution digital elevation models, *ISPRS Journal of Photogrammetry and Remote Sensing* **50**: 38 – 49.

Whelan, P. F. and Molloy, D. (2001): *Machine Vision Algorithms in Java: Techniques and Implementation*, Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Wikipedia (2007): Minimum bounding rectangle — wikipedia, the free encyclopedia: Online; Last access 19 March 2007.

Wikipedia (2008): Outlier — wikipedia, the free encyclopedia: Online; Last access May 2008.

Wikipedia (2009): Scale-invariant feature transform — wikipedia, the free encyclopedia: Online; Last access 10 October 2008.

Zhan, Q., Molenaar, M. and Tempfli, K. (2002): Building extraction from laser data by reasoning on image segements in elevation slices, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume 34 (B3), 305–308.

Zhang, K., Chen, S., Whitman, D., Shyu, M., Yan, J. and Zhang, C. (2003): A progressive morphological filter for removing non-ground measurements from airborne lidar data, *IEEE Transaction on Geoscience and Remote Sensing* **41**(4): 872–882.