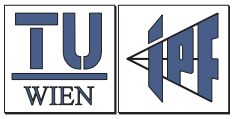
GEOWISSENSCHAFTLICHE MITTEILUNGEN

Heft Nr. 83, 2008

AUTOMATIC EXTRACTION OF BUILDINGS FROM AIRBORNE LASERSCANNER DATA AND AERIAL IMAGES

von
Franz Rottensteiner



Veröffentlichung des Institutes für
Photogrammetrie und Fernerkundung

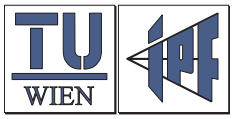
**Schriftenreihe der Studienrichtung VERMESSUNG UND GEOINFORMATION
TECHNISCHE UNIVERSITÄT WIEN**

GEOWISSENSCHAFTLICHE MITTEILUNGEN

Heft Nr. 83, 2008

AUTOMATIC EXTRACTION OF BUILDINGS FROM AIRBORNE LASERSCANNER DATA AND AERIAL IMAGES

von
Franz Rottensteiner



Veröffentlichung des Institutes für
Photogrammetrie und Fernerkundung

**Schriftenreihe der Studienrichtung VERMESSUNG UND GEOINFORMATION
TECHNISCHE UNIVERSITÄT WIEN**

Herausgeber und Verleger: Univ.-Prof. Dr. Wolfgang Wagner
Vorstand des Institutes für Photogrammetrie und Fernerkundung
der Technischen Universität Wien
A-1040 Wien, Gußhausstraße 27-29

Die Kosten für den Druck wurden aus eigenen Einnahmen des Institutes für Photogrammetrie und Fernerkundung der Technischen Universität Wien getragen.

Diese Arbeit wurde an der Fakultät für Mathematik und Geoinformation der Technischen Universität Wien (Karlsplatz 13, A-1040 Wien, Österreich) als Habilitationsschrift eingereicht.

Gutachter:

Ao. Univ.-Prof. Dr. Margrit Gelautz	Institut für Softwaretechnik und Interaktive Systeme der Technischen Universität Wien.
Univ.-Prof. Dr. Helmut Pottmann	Institut für diskrete Mathematik und Geometrie der Technischen Universität Wien.
Prof. Dr.-Ing. Christian Heipke	Institut für Photogrammetrie und GeoInformation der Leibniz Universität Hannover.
Prof. Dr. George Vosselman	International Institute for Geo-Information Science and Earth Observation (ITC), Enschede.

Tag des Habilitationskolloquiums: 24. 11. 2008

Druck: Grafisches Zentrum HTU GmbH, A-1040 Wien
Auflage: 100 Stück
ISSN 1811-8380

Abstract

This habilitation presents a collection of papers dealing with the automatic extraction of buildings from Airborne Laserscanner (ALS) data, supported by aerial imagery. Building extraction consists of two stages: the detection of buildings, essentially a classification task, and the geometrical reconstruction of buildings in previously detected regions of interest. Both stages are dealt with in this work.

First, a rule-based method for building detection is presented. This method can use both ALS data and multi-spectral information in the form of a normalised difference vegetation index (NDVI). This method can be applied in a hierarchical framework of coarse generation of a digital terrain model by morphological filtering. The second method for building detection presented in this work is based on the Dempster-Shafer theory for data fusion. It uses a heuristic model for the distribution of evidence to the classes of the classification process. A thorough evaluation of that method has shown that this model is appropriate and that most of its parameters can be determined relatively easily from "meaningful" entities such as a minimal building height or the approximate percentage of trees in a scene. It was shown that buildings larger than about 120 m^2 can be reliably detected using ALS data of a resolution of 1 m and an NDVI image. Buildings between 50 m^2 and 120 m^2 can still largely be detected. The major influence of the NDVI was a reduction of false positive detections of buildings smaller than 100 m^2 by up to 15%.

Building reconstruction as presented in this work starts with the extraction of roof planes from the ALS data. After that, a classification of the mutual geometrical relations between neighbouring roof planes is carried out, with the aim of determining the boundary polygons of these roof planes. This includes a method for the precise location of step edges in ALS data. In this process, decisions are based on statistical tests rather than on simple thresholding operations, thus increasing the robustness of the approach. These tests require rigorous modelling of the stochastic properties of the geometric entities involved. The roof boundary polygons can be grouped to form polyhedral building models. Finally, the parameters of these polyhedral models are estimated in a consistent parameter estimation process that considers geometrical regularities. In this way, building models with a planimetric accuracy in the range of the original point spacing and with a height accuracy in the range of a few centimetres can be generated. However, the quality of the results is limited by the sensor resolution, since the planar segmentation requires a certain minimum number of ALS points on each plane of the roof.

Kurzfassung

In dieser Habilitationsschrift sind mehrere wissenschaftliche Arbeiten zusammengefasst, die sich mit der automatischen Extraktion von Gebäuden aus flugzeuggestützten Laserscannerdaten befassen. Der Prozess der Gebäudeextraktion umfasst dabei zwei Arbeitsschritte: die Detektion von Gebäuden, im Wesentlichen eine Klassifizierungsaufgabe, und die geometrische Rekonstruktion von Gebäuden in den zuvor detektierten Interessensgebieten. Beide Themen werden in dieser Arbeit behandelt.

Zunächst wird eine regelbasierte Methode zur Gebäudedetektion präsentiert. Diese Methode kann sowohl Laserscannerdaten als auch multispektrale Information über den "normalised difference vegetation index" (NDVI) berücksichtigen. Sie kann auch in einem hierarchischen Kontext zur Erzeugung eines digitalen Geländemodells durch morphologische Filterung angewandt werden. Die zweite Methode zur Gebäudedetektion, die in dieser Arbeit vorgestellt wird, beruht auf der Theorie von Dempster-Shafer zur Datenfusion. In dieser Methode wird ein heuristisches Modell für die Verteilung der in den Daten enthaltenen Evidenz bezüglich der im Klassifizierungsprozess zu unterscheidenden Klassen verwendet. Eine sorgfältige Evaluierung dieser Methode hat gezeigt, dass dieses Modell gut für die Klassifizierungsaufgabe geeignet ist und dass die meisten Parameter dieses Modelles auf relativ einfache Weise aus "sinnvollen" Größen wie z. B. der minimalen Gebäudehöhe oder einem geschätzten Prozentsatz von Bäumen in der Szene abgeleitet werden können. Es konnte weiters gezeigt werden, dass Gebäude größer als ca. 120 m^2 zuverlässig aus Laserscannerdaten mit einer Auflösungvon etwa 1 m und einem NDVI-Bild abgeleitet werden können. Gebäude mit einer Fläche zwischen 50 m^2 und 120 m^2 können ebenfalls weitgehend detektiert werden. Der wesentliche Einfluss des NDVI war eine Reduktion der Anzahl der fälschlich detektierten Gebäude mit einer Fläche kleiner als 100 m^2 um bis zu 15%.

Die Rekonstruktion von Gebäuden, wie sie in dieser Arbeit verstanden wird, beginnt mit der Extraktion von Dachflächen aus den Laserscannerdaten. Danach wird eine Klassifizierung der geometrischen Relationen zwischen benachbarten Dachflächen durchgeführt, die das Ziel hat, die Randpolygone dieser Dachflächen zu bestimmen. Dies inkludiert eine Methode für die genaue Lokalisierung von Höhensprüngen in den Laserscannerdaten. In diesem Prozess werden Entscheidungen auf Grund von statistischen Tests getroffen anstatt auf der Basis von einfachen Schwellwertbildungen, wodurch die Robustheit der Methode erhöht wird. Um diese statistischen Tests durchführen zu können, ist eine strenge Modellierung der stochastischen Eigenschaften der involvierten geometrischen Grössen von Nöten. Die Dachrandpolygone können zu Polyedermodellen gruppiert werden. Schlieẅslich werden die Parameter dieser Polyedermodelle auf konsistente Weise geschätzt, wobei im Schätzprozess auch geometrische Bedingungen zur Regularisierung der Modelle berücksichtigt werden. Auf diese Weise können Gebäudemodelle mit einer Lagegenauigkeit erzeugt werden, die in etwa dem ursprünglichen Punktabstand entspricht, während ihre Höhengenauigkeit im Bereich von wenigen Zentimetern liegt. Allerdings wird die Qualität der Ergebnisse stark durch die

Sensor-Auflösung limitiert, weil die Segmentierung der Dachebenen eine gewisse minimale Anzahl von Laserpunkten in jeder Dachebene benötigt.

Contents

Chapter 1

Introduction

Airborne laserscanning (ALS), also referred to as LiDAR (*Li*ght *D*etection *A*nd *R*anging), has proven itself to be a powerful technique for data acquisition for Geographic Information Systems (GIS). It essentially delivers a 3D point cloud, as opposed to a 2D image provided by an aerial camera. It thus not only gives direct access to the third dimension, but via an analysis of the local geometry of the point cloud also provides information about surface parameters such as surface roughness. These advantages of ALS data compared to aerial imagery are contrasted by a certain lack of resolution and by the lack of multi-spectral information, which makes the fusion of these data sources an interesting option for data acquisition for GIS. In this context, the automation of topographic object extraction has been an important topic of research in photogrammetry, remote sensing, and computer vision for about two decades. Among the objects that are to be mapped automatically, man-made structures such as buildings and roads are of utmost importance. Despite the enormous efforts spent in the past, progress can only be characterised as incremental.

This habilitiation presents the author's scientific work in the field of building extraction from ALS data, along with some contributions to the fusion of ALS data and aerial imagery for that purpose. First, a description of the major stages of building extraction will be given in section 1.1, followed by a description of how the individual chapters contribute to this process that will be presented in section 1.2. Except for this introductory chapter, each of the following chapters corresponds to a scientific paper that has been previously published either in a scientific journal or in conference proceedings. All these papers were subject to a full-paper peer-reviewing process:

- **Chapter 2:** Franz Rottensteiner and Christian Briese: A new method for building extraction in urban areas from high-resolution LIDAR data. In: *Proceedings of the ISPRS Commission III Symposium in Graz, Austria, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. XXXIV - 3A, pp. 295-301, 2002*. Peer-reviewed conference paper, cited in the text as [Rottensteiner and Briese, 2002].

- **Chapter 3:** Franz Rottensteiner, John Trinder, Simon Clode, and Kurt Kubik: Building detection using LIDAR data and multispectral images. In: *Proceedings of the APRS Conference on Digital Image Computing: Techniques and Applications (DICTA), Sydney, Australia, Vol. II, pp. 673-682, 2003*. Peer-reviewed conference paper, cited in the text as [Rottensteiner et al., 2003]

- **Chapter 4:** Franz Rottensteiner, John Trinder, Simon Clode, and Kurt Kubik: Using the Dempster Shafer method for the fusion of LIDAR data and multi-spectral images for building detec-

tion. In: *Information Fusion 6 (4), pp. 283-300, 2005*. Peer-reviewed journal paper, cited in the text as [Rottensteiner et al., 2005b].

- **Chapter 5:** Franz Rottensteiner, John Trinder, Simon Clode, and Kurt Kubik: Building detection by fusion of airborne laserscanner data and multi-spectral images: Performance evaluation and sensitivity analysis. In: *ISPRS Journal of Photogrammetry and Remote Sensing 62(2), pp. 135-149, 2007*. Peer-reviewed journal paper, cited in the text as [Rottensteiner et al., 2007].
- **Chapter 6:** Franz Rottensteiner: Automatic generation of high-quality building models from LIDAR data. In: *IEEE Computer Graphics and Applications 23(6), pp. 42-51, 2003*. Peer-reviewed journal paper, cited in the text as [Rottensteiner, 2003].
- **Chapter 7:** Franz Rottensteiner, John Trinder, Simon Clode, and Kurt Kubik: Automated delineation of roof planes in LIDAR data. In: *Proceedings of the ISPRS workshop on Laserscanning in Enschede, The Netherlands, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. XXXIV-3/W19, pp. 221-226, 2005*. Peer-reviewed conference paper, cited in the text as [Rottensteiner et al., 2005a].
- **Chapter 8:** Franz Rottensteiner: Consistent estimation of building parameters considering geometric regularities by soft constraints. In: *Proceedings of the ISPRS Commission III Symposium in Bonn, Germany, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. XXXIV - 3, pp. 13-18, 2006*. Peer-reviewed conference paper, cited in the text as [Rottensteiner, 2006].

All these papers are reproduced in the way they originally appeared, with a few exceptions:

- They are all formatted in a consistent way.
- Sections, equations, tables, and figures were numbered in a consistent way throughout the text. Also, the spelling of references to equations, tables, and figures was homogenized.
- The scaling of some of the figures was changed.
- The chapters are referred to as "chapter" (rather than "paper" in the original papers).
- Some minor typographic errors were corrected.
- A few footnotes were added.

1.1 Stages of Building Extraction

Building extraction as it is understood in the context of this work consists of three stages:

1. **Pre-processing:** This stage comprises preparatory work, including aerial triangulation for the orientation of aerial images, bias correction of the ALS point cloud, generation of orthophotos if necessary, and the estimation of height grids from the ALS points. The latter is required because with a few exceptions, all algorithms described in this work operate on a *Digital Surface Model (DSM)* in the form of a height grid. Pre-processing does not constitute a part of this work. Throughout this work, images are thus assumed to be oriented and/or geocoded, ALS data are considered to be aligned with image data with sufficient accuracy, and DSM grids are supposed to exist.

2. **Building Detection:** This is essentially a classification of the original sensor data in order to recognise buildings and to locate them in an approximate way. It is often divided into three interrelated steps:

   (a) *Generation of a Digital Terrain Model (DTM):* A DTM is a model of the earth surface in the sense of the "bare earth", without vegetation, buildings, and other objects. Sometimes, DTM generation is considered an integral part of building detection, because it is also essentially a classification (terrain points vs. off-terrain points). DTM generation is necessary because the height of a point above the terrain is an important classification cue for building detection.

   (b) *Detection of building candidate regions:* In this phase, each ALS point or each grid point of the DSM is classified independently according to whether it corresponds to a building or to another object class. Connected segments of "building points" or "building pixels" are considered to be initial building regions. The most important problem to be tackled in this context is the separation of buildings and trees, especially if trees grow in the vicinity of buildings.

   (c) *Final classification of building candidate regions:* A second classification process is applied to the initial building regions to eliminate regions that actually correspond to other objects, mostly trees.

   In many industrialised countries, building data do already exist in national data bases such as the cadastre. Often these existing data are only available in 2D, but they provide precise (if not necessarily up-to-date) information about where a building exists. If such data are available, building detection can either be neglected, or it can be used to first detect changes that have occurred in the period between the collection of the 2D building data and the acquisition of the sensor data used for building extraction. However, for the algorithms presented in this work, no existing building data are required.

3. **Building Reconstruction:** Building detection essentially delivers coarse building outlines, e.g. represented as closed (2D) polygons. In the third stage of building extraction, buildings have to be geometrically reconstructed. The outcome of building reconstruction as it is understood in this work should be 3D models of the buildings that correctly resemble the roof structure, but with some degree of generalisation that is caused by the limitations of the sensor resolution. As will be pointed out several times in the subsequent chapters, there are different ways of modelling buildings and of representing the buildings in the reconstruction process. In this work, buildings are modelled by polyhedrons having vertical walls and no roof overhangs, and a bottom-up strategy for building reconstruction consisting of three major stages is applied:

   (a) *Detection of roof planes:* As stated above, ALS data give direct access to surface roughness parameters. That means that it is possible to detect connected regions in the data that can geometrically be described by a plane. As buildings mostly have planar roof surfaces, such planar segmentation algorithms can be applied to detect roof planes. In this work, planar segmentation will be applied to DSM grids, and it will be based on local co-planarity of grid points; alternative approaches capable of handling the original ALS point clouds by clustering techniques have been proposed in [Peternell and Steiner, 2004] and [Pottmann et al., 2002].

   (b) *Grouping and model generation:* The roof planes have to be combined to consistent 3D building models. This involves the determination of the roof boundary polygons and, along

with that, of the mutual topological relations between the roof planes. As soon as these topological relations are known the roof planes can be grouped.

(c) *Consistent estimation of the model parameters and geometric regularisation:* The parameters of the resulting models have to be estimated in a consistent way from all the available sensor data. This problem is often overlooked, but is still essential for the generation of high-quality building models. Regularisation means that geometrical constraints are introduced where evidence supporting such a hypothesis is found. In the papers collected in this work, regularisation is sometimes considered a separate stage of building reconstruction. Regularisation is carried out using an expansion of the method for representing the geometric structure of the polyhedral models in the parameter estimation process.

1.2 Contributions of the Individual Chapters to the Problem of Building Extraction

Chapter 2 was the first of the papers collected in this work to be published. It is the only chapter in this work contributing to both building detection and reconstruction:

1. It presents a new method for DTM generation from ALS data in densely built-up areas that is based on a hierarchical application of robust linear prediction[1].

2. Based on the DTM thus generated, rule-based classification algorithms are applied for the generation of initial building regions and for the final classification of building regions. In this context, a specific way of representing surface roughness in grid-based DSMs based on the Förstner operator for the extraction of features from digital images was applied for the first time. This chapter also presents a technique to eliminate trees adjacent to buildings. The method is applied to a relatively complex scene. A coarse visual check of the results is carried out, and the results are found to be satisfactory.

3. A new technique for roof plane detection is presented that is based on the same method for representing surface roughness as building detection. The subsequent stages of building reconstruction are outlined but not yet tackled in this chapter. Preliminary results of roof plane extraction are found to be encouraging, but no evaluation beyond a visual inspection is carried out.

Chapter 3 deals with building detection only. The rule-based method introduced in chapter 2 is modified and expanded in several ways:

1. The rule-based classification method for building detection is expanded so that it can also use a *Normalised Difference Vegetation Index (NDVI)*.

2. A new hierarchical framework for building detection is introduced that no longer relies on hierarchic robust estimation for DTM generation, but applies morphologic filtering for that purpose. This framework considers the fact that building detection and DTM generation are dependent on each other: in order to reliably determine a DTM, large buildings have to be detected (and eliminated) first, whereas for building detection, a coarse DTM is essential.

[1] This is the contribution of my co-author Christian Briese

3. Experiments are presented for test site of suburban characteristics, with more heterogeneous types of development than the test site used in chapter 2. Again, no evaluation beyond a coarse visual inspection is carried out. However, it is acknowledged that some of the thresholds for the rule-based classification are not easily tuned and that for the initial classification, a probabilistic method evaluating all sensor data at the same time might be a better choice for detecting smaller buildings.

Chapter 4 introduces such a probabilistic classification method for building detection. Its contributions to the overall problem are as follows:

1. It contains a thorough literature review on building detection.

2. The model for surface roughness used in all chapters throughout this work is explained in detail.

3. The rule-based method for building detection described in chapter 3 is only used for detecting the largest buildings in the process of hierarchical DTM generation.

4. The chapter presents a new classification method for building detection based on Dempster-Shafer fusion of ALS data and multi-spectral imagery. In this context, heuristic models for the stochastic properties of the sensor data are developed, using the important property of Dempster-Shafer fusion that it can easily model insufficient knowledge about the a priori distributions of the sensor data with respect to the classes to be discerned.

5. A thorough evaluation of the classification results is carried out. In this context, a methodology for evaluation capable of comparing two data sets of different topology is introduced. The results show that buildings larger than 90 m^2 could be reliably detected in data of a resolution of about 1.2 m.

Chapter 5 is the final chapter on building detection in this work. Its contributions are as follows:

1. It presents the algorithm described in chapter 4 in an improved form, especially with respect to the way surface roughness is modelled. It also presents a new way of considering the uncertainty of the NDVI in shadow areas.

2. It validates the statistical models used in the Dempster-Shafer classification process using two data sets of different sensor and scene characteristics. In this context, rules for tuning the parameters of the statistical model are discussed. The new model for surface roughness helps to replace a parameter that cannot be chosen in an intuitive way by the user by a very "intuitive" parameter (an estimate for the percentage of trees in the scene).

3. The evaluation of the method is expanded in order not only to provide results for optimal parameters, but also to include a sensitivity analysis to determine the influence of parameter tuning on the quality of the results. The conclusion is that reasonably good results can be achieved with parameter settings that are not optimal, but in order to exploit the full potential of the model for buildings that are relatively small compared to the sensor resolution, the parameters have to be tuned with care.

4. In the evaluation of the method, the contributions of the individual cues to the quality of the classification results are assessed. It is shown that, whereas the multi-spectral data do not help much in the classification of large building structures, they can improve the correctness of the results for smaller buildings by up to 15%.

Chapter 6 contributes to the topic of building reconstruction as follows:

1. It expands the method for building roof detection originally presented in chapter 2 by an iterative scheme designed to avoid the selection of a single threshold for the classification of planar surfaces, giving a statistical interpretation for the procedure.
2. It describes a strategy for the delineation of roof polygons, presenting first results. It is concluded that an improved method for the classification of the mutual relations between neighbouring roof planes is required.
3. It presents a generic model for the consistent estimation of building parameters, which at the time of writing had not yet been fully implemented.
4. It presents ideas to the topic of data fusion for building reconstruction, aiming at establishing a "multi-sensor-grammetry". These ideas were later worked out in [Rottensteiner et al., 2004], but this publication is not included here because it was not peer-reviewed.

Chapter 7 is a follow-up on the method presented in chapter 6. Its focus is on the second stage of building reconstruction, i.e. the delineation of the roof boundary polygons and grouping of roof planes:

1. It describes a unique way of classifying the mutual relations between neighbouring roof planes as either being intersecting in a consistent way or not intersecting, thus indicating a step edge.
2. It describes a new method for the precise location of step edges corresponding to walls in ALS data. This method makes use of specific knowledge about the appearance of buildings in these data to avoid false step edge elements at trees that are adjacent to buildings.
3. It describes a method for generating consistent roof boundary polygons, detecting situations where neigbouring roof boundary polygons do not intersect in a consistent way, thus indicating short missed step edge segments.
4. In all these processes, decisions that have to be taken are based on statistical tests and robust estimation rather than on simple thresholding of discrepancies. For this purpose, the concept of uncertain projective geometry as it was proposed by [Heuel, 2004] is applied to geometric reasoning on geometric primitives derived from ALS data. In this way, the selection of user-defined thresholds is kept to a minimum and largely replaced by the selection of a significance level for the statistical tests.
5. Examples are given to show how the algorithm can cope with rather complex roof shapes. The adjustment model first proposed in chapter 6 is applied manually to a few buildings, which results in numbers for the precision of the building vertices thus derived.

Chapter 8 deals with the final stage of building reconstruction, i.e. with the consistent estimation of building parameters and with the inclusion of geometric constraints:

1. It gives a systematic overview on how geometric regularities can be considered in building extraction.

2. The method for parameter estimation considering geometric regularities by ”soft constraints” originally presented in chapter 6 is revised and presented in a generalised way.

3. It is shown how this method can be applied to different scenarios of building reconstruction.

4. An example is worked out in order to quantify the accuracy that can be achieved by the building reconstruction method described in chapters 2 and 6 to 8. This evaluation shows that relatively complex roof structures can be reconstructed using the method for roof plane delineation described in chapter 7, with accuracies typically in the order of the point spacing in planimetry and in the range of a few centimetres in height. The dependency of these numbers on the size of the roof plane is evaluated, and it is shown that in most cases, the overall adjustment including geometric constraints improve the accuracy, especially in planimetry. The example also shows the limitations of the method: roof planes receiving too few laser strikes cannot be detected, and step edges with a small height difference between the neighbouring roof planes cannot be located precisely.

Chapter 2

A New Method for Building Extraction in Urban Areas from High-resolution LIDAR Data

2.1 Introduction

2.1.1 Motivation and Goals

Automation in data acquisition for 3D city models is an important topic of research with the goal of reducing the costs of providing these data at an appropriate level of detail. In addition to photogrammetric techniques relying on aerial images, the generation of 3D building models from point clouds provided by LIDAR sensors is gaining importance. This development has been triggered by the progress in sensor technology which has rendered possible the acquisition of very dense point clouds using airborne laser scanners. Using LIDAR data with point densities of up to one point per square meter, it is possible not only to detect buildings and their approximate outlines, but also to extract planar roof faces and, thus, to create models which correctly resemble the roof structures.

Building extraction is solved in two steps [Brenner, 2000]. First, buildings have to be detected in the data, and the approximate building outlines have to be determined. Second, in the regions of interest thus detected, the buildings have to be reconstructed geometrically, which results in 3D polyhedral models of the buildings. It is the goal of this chapter to present a new method for the automatic creation of polyhedral building models in densely built-up areas from high-resolution LIDAR data without using ground plans. Our method is unique with respect to the algorithms used for building detection because it is based on robust interpolation. In the detected building regions, planar roof patches, their bounding polygons, and their neighbourhood relations are extracted. Grouping of neighbouring planes has not yet been implemented. The examples presented in this chapter were computed using the LIDAR data from a test site in the City of Vienna captured by TopoSys. The resolution of the original point cloud is 0.1 m (in-flight) by 1 m (cross-flight). A grid of 0.5×0.5 m^2 derived from that point cloud was used for building extraction. The test data were captured in the course of a pilot project for the Municipality of Vienna in order to evaluate and compare various techniques for the generation of 3D city models. Our intermediate results show the high potential of the method presented in this chapter.

2.1.2 Related Work

There have been several attempts to detect buildings in LIDAR data in the past. The task has been solved by classifying the LIDAR points according to whether they belong to the terrain, to buildings or to other object classes, e.g., vegetation. Morphological opening filters or rank filters are commonly used to determine a digital terrain model (DTM) which is subtracted from the digital surface model (DSM). By applying height thresholds to the normalized DSM thus created, an initial building mask is obtained [Weidner, 1997, Ameri, 2000a]. The initial classification has to be improved in order to remove vegetation areas. In [Brunn and Weidner, 1997], this is accomplished by a framework for combining various shape cues in a Bayesian network. Our algorithm for building detection from LIDAR points is based on the method for DTM generation by robust interpolation presented in [Kraus and Pfeifer, 1998].

The geometrical reconstruction of the buildings in previously detected regions of interest has been tackled in two ways. First, parametric primitives can be instantiated and fitted to the data if sufficient evidence is found. Second, planar patches can be detected in a DSM created from the LIDAR points, and polyhedral building models can be derived by grouping these planar patches. As parametric primitives often have a rectangular footprint, they are especially used if 2D ground plans giving a precise location of the building outlines are available. The polygon delineating a building in a 2D map is split into rectangular regions. In each rectangle, the parameters of parametric models are determined using the DSM, and the model achieving the best fit is accepted [Brenner, 2000, Vosselman and Dijkman, 2001]. The data driven generation of polyhedral building models from LIDAR data only makes sense if the point density is high enough so that a sufficient number of data points is located at least in the most relevant planes of the roofs. As the building outlines are difficult to be located precisely, again ground plans are often used for that purpose. Ground plans also reduce search space for the estimation of the parameters of adjoining planar patches because the gradient direction of such planes is usually perpendicular to the adjacent polygon segment in the ground plan [Haala et al., 1998, Brenner, 2000, Vosselman and Dijkman, 2001]. Initial planar patches are found by a segmentation of the DSM. [Brenner, 2000] gives several methods for DSM segmentation, e.g., the analysis of surface curvature, i.e., of changes in the surface normal vectors, or a segmentation taking into account the directions of the polygon segments of a ground plan.

As soon as the initial planar patches have been found, neighbouring patches are grouped [Baillard et al., 1999], and the polygons delineating the borders of planar patches have to be found. The latter task involves finding consistent intersections at the building vertices [Moons et al., 1998]. Finally, the 3D border polygons have to be combined in order to obtain consistent building models. At the building outlines, vertical walls, and, finally, the floor have to be added to the model. A coarse-to-fine strategy can be applied by first searching for the most relevant structures in the data and using refined methods for modelling the buildings in regions not being "explained" sufficiently by the initial models [Vosselman and Dijkman, 2001]. The problem of precisely determining the building outlines has been tackled by [Weidner, 1997] by applying the minimum description length principle for deciding on regularisations.

2.2 Work Flow for Building Extraction

The work flow for the extraction of buildings from LIDAR points is presented in figure 2.1. The first step is the interpolation of a DSM and a DTM from the original data at an appropriate resolution. Our

method for DTM generation which performs a classification of the original points into terrain versus off-terrain points by robust estimation will be explained in section 2.3. From this instance onwards, the models created by interpolation are used, no longer the original data points.

By subtracting the DTM from the DSM and by applying a threshold to the height differences, an initial building mask is created which still contains vegetation and other objects. Binary morphological operators and an analysis of the DSM texture, i.e., of the local variations of the DSM normal vectors, are used to eliminate these areas. The final results of building detection, i.e., the individual building regions, are found by a connected component analysis. Building detection by comparing the DTM and the DSM is described in section 2.4.

In the building candidate regions, a plane segmentation based on an analysis of the variations of the DSM normal vectors is applied to find planar patches. These patches are expanded by region growing algorithms. In the current version, the neighbourhood relations of these patches are determined, and a simple model resembling the roof structure of the building is created. In the future, neighbouring planes will be grouped consistently before the initial building models containing the most relevant roof structures are created. In a post-processing phase, the model has to be refined in order to contain details originally not detected. The current state of our technique for geometrical reconstruction of roof structures from a DSM is described in section 2.5.

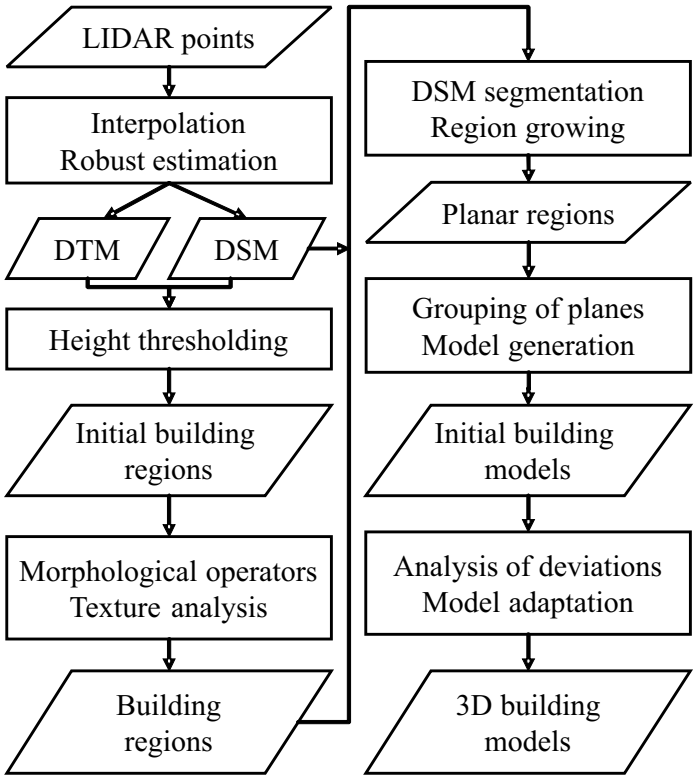


Figure 2.1: Work flow for building extraction from LIDAR data.

2.3 DTM Generation from LIDAR Data in Urban Regions

An algorithm for the automatic generation of DTMs in forested regions from laser scanner data was developed at our institute. This method is based on iterative robust interpolation of a DTM grid, and it combines the elimination of off-terrain points and the interpolation of the DTM grid in one process [Kraus and Pfeifer, 1998]. For the generation of a DTM in densely built-up areas, this method has to be modified to work in a hierarchical framework [Briese et al., 2001]. With this coarse-to-fine approach it is possible to cope with relatively large areas without terrain points (e.g., large building blocks).

2.3.1 Robust Interpolation

In an iterative process the irregularly distributed LIDAR points are weighted in a way that the modelled surface describes the terrain. The classification of the points in terrain versus off-terrain points is performed by thresholding the discrepancies to the computed surface by user-specified tolerance values.

In a first step, a coarse approximation of the surface is computed taking into account all available LIDAR points. Next, the discrepancies, i.e., the differences of the heights of the LIDAR points and the interpolated surface at the planimetric positions of the LIDAR points, are computed. The discrepancies are the parameters of a weight function which is used to assign an individual weight to each point in the subsequent processes. The interpolation of the DTM is repeated, the weights of the LIDAR points being modulated depending on the discrepancies of the most recent iteration. This iterative process is terminated as soon as a stable situation or a maximum number of iterations is reached. Two types of models are used in our algorithm, i.e., the functional model which defines the way the surface is computed, and the stochastic model which is responsible for weighting.

The functional model: Linear prediction is used for modelling the surface. Using this model, it is possible to compute a smooth surface considering random measurement errors [Kraus, 2000].

The stochastic model: For the generation of a DTM, high weights must be assigned to terrain points below or on the averaging surface, and low weights have to be assigned to the non-terrain points which are above the averaging surface. A typical weight function $p(r)$ parameterized by the discrepancies r for the generation of a DTM from laser scanner data is presented in figure 2.2. The weight function we use is not symmetrical, and it is shifted by a value g. It has a sharp decline defined by its half-width value h and slant s for discrepancies greater than its central point (i.e., for off-terrain points above the estimated surface) and no decline for the terrain points. The exclusion of points from the interpolation process is triggered by a threshold t derived from a user-specified tolerance for the size of the discrepancies. For a comprehensive description of this algorithm see [Kraus and Pfeifer, 1998].

2.3.2 Hierarchic Robust Interpolation

The method of iterative robust interpolation relies on a "good mixture" of terrain and off-terrain points. Therefore, this algorithm does not work in large areas without terrain points as they are likely to exist in densely built-up areas. To provide this "good mixture" also in densely built-up

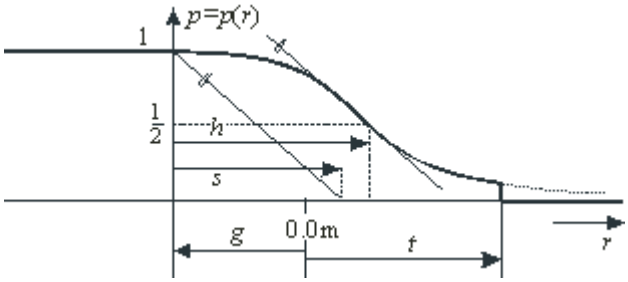


Figure 2.2: Weight function for the generation of a DTM from laser scanner data.

areas, robust interpolation has to be applied in a hierarchic way using data pyramids (comparable to image pyramids in image processing). The hierarchic robust interpolation proceeds as follows:

1. Create the data pyramids. This can be achieved by selecting, for instance, the lowest points in a regular grid mesh.
2. Perform robust interpolation to generate a DTM.
3. Compare the DTM to the data of the next higher resolution and accept points within a certain tolerance band.

Steps 2 and 3 are repeated at each resolution level of the data pyramid. The results of DTM interpolation in the lower resolution levels are used for the computation of the surface in the next higher resolution because only points having passed the thresholding step 3 are considered at that level.

In [Briese, 2000], this strategy has been evaluated for the generation of a high-quality DTM of a test site located in the City of Vienna (2.5 km^2) using three data pyramid levels (5 m, 2 m and 0.5 m). A few intermediate results of this DTM generation process are presented in the perspective views in figures 2.3 and 2.4.

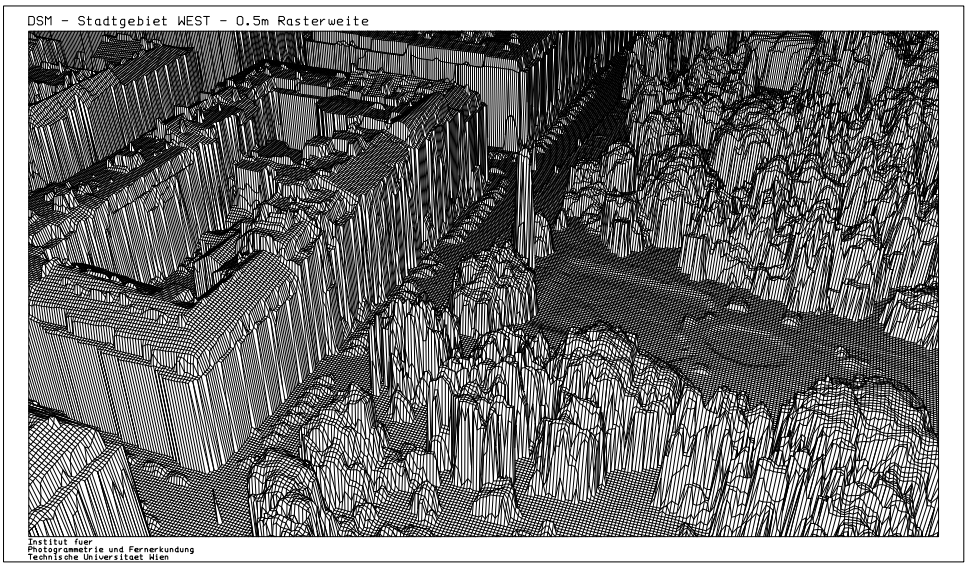
Figure 2.5 shows a perspective view of a detail of the final DTM. Further details about hierarchical robust interpolation, its implementation in the software package SCOP, and the results of some further examples can be found in [Briese et al., 2001].

2.4 Building Detection

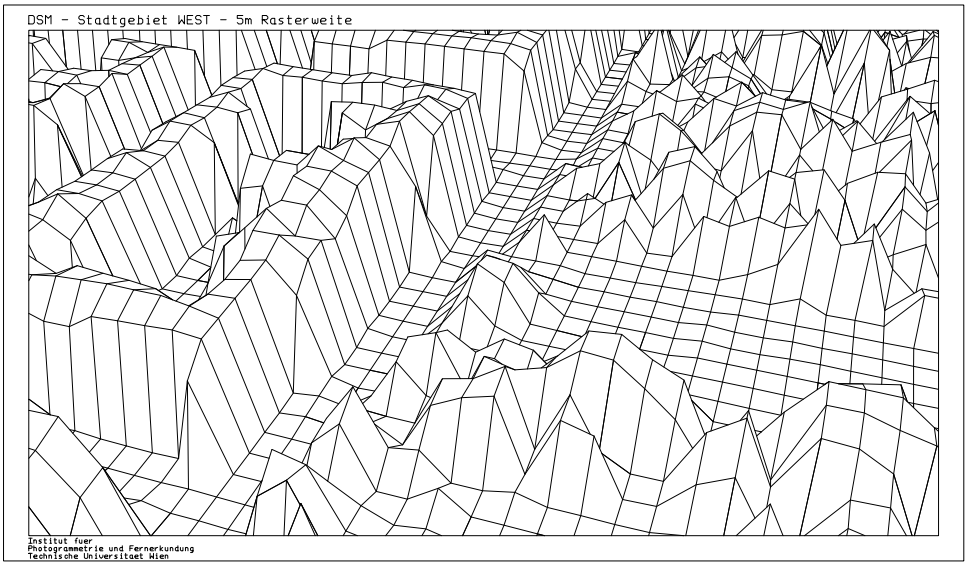
As described in section 2.3, two digital elevation models are derived by interpolation: a DTM is computed from the points classified as "terrain points" with a high degree of smoothing, whereas a DSM is computed from all points without smoothing (figure 2.6(a)). An initial building mask is created by thresholding the height differences between the DSM and the DTM (e.g., by h_{min} = 3.5 m). This initial building mask still contains areas covered by vegetation, and some individual building blocks are not correctly separated (figure 2.6(b)). A morphological opening filter using a small (e.g., 5×5) square structural element is applied to the initial building mask in order to erase small elongated objects such as fences and to separate regions just bridged by a thin line of pixels. A connected component analysis of the resulting image is applied to obtain the initial building regions. At this

instance, regions smaller than a minimum area (e.g., 40 m^2) and regions at the border of the DSM are discarded (figure 2.6(c)).

Some of the remaining regions in figure 2.6(c) still correspond to groups of trees. These regions can be eliminated by evaluating a "terrain roughness" criterion derived by an analysis of the second derivatives of the DSM. In [Fuchs, 1998], a method for polymorphic feature extraction is described which aims at a classification of texture as being homogeneous, linear, or point-like, by an analysis of the first derivatives of a digital image. This method is applied to the first derivatives of the DSM using a large (e.g., 9×9) integration kernel. For each initial building region, the number of "point-like" pixels is counted. Regions containing more than 50% of pixels classified as being "point-like"

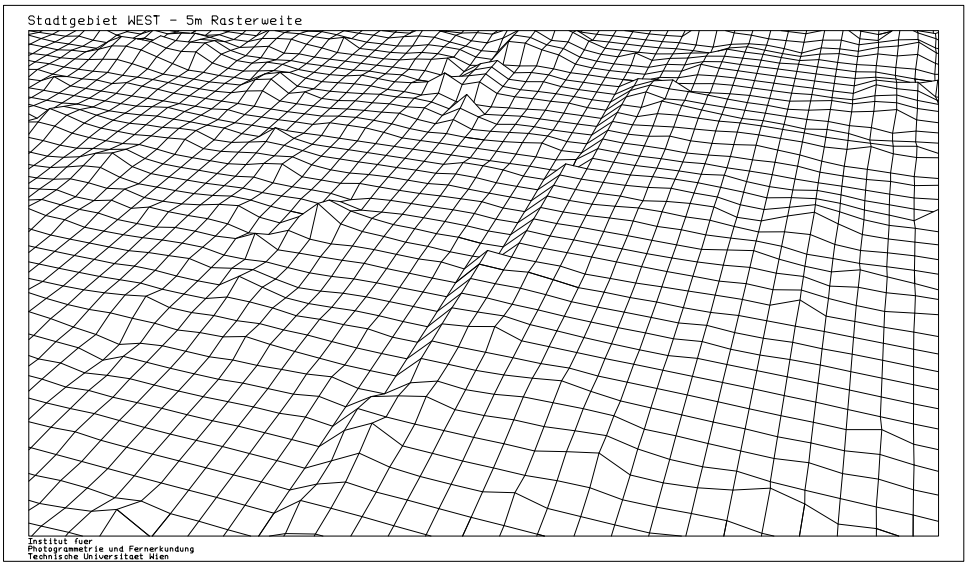


(a)

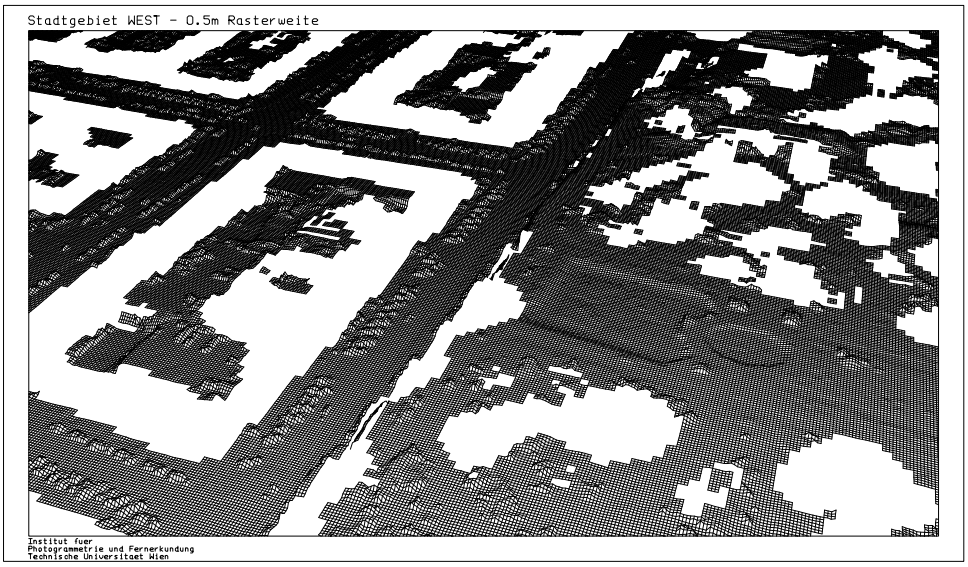


(b)

Figure 2.3: (a) DSM computed from all LIDAR points (0.5 m). (b) DSM at the lowest resolution (5 m).



(a)



(b)

Figure 2.4: (a) DTM at the lowest resolution (5 m). (b) DSM (0.5 m) of the accepted original points (the points within a user-defined tolerance band).

(thus, pixels being in a neighbourhood of great, but anisotropic variations of the surface normals) are very likely to contain vegetation rather than buildings, and they are eliminated. Figure 2.6(d) shows the results of texture classification. Note the obvious co-incidence of clusters of "point-like" pixels displayed in black and vegetation areas such as those in the botanical gardens on the left margin of the test site.

The terrain roughness criterion is very efficient in classifying isolated vegetation regions, but it cannot find vegetation areas which are still connected to buildings. In a final stage of analysis, we try to eliminate such areas. By morphological opening using a square structural element, regions just

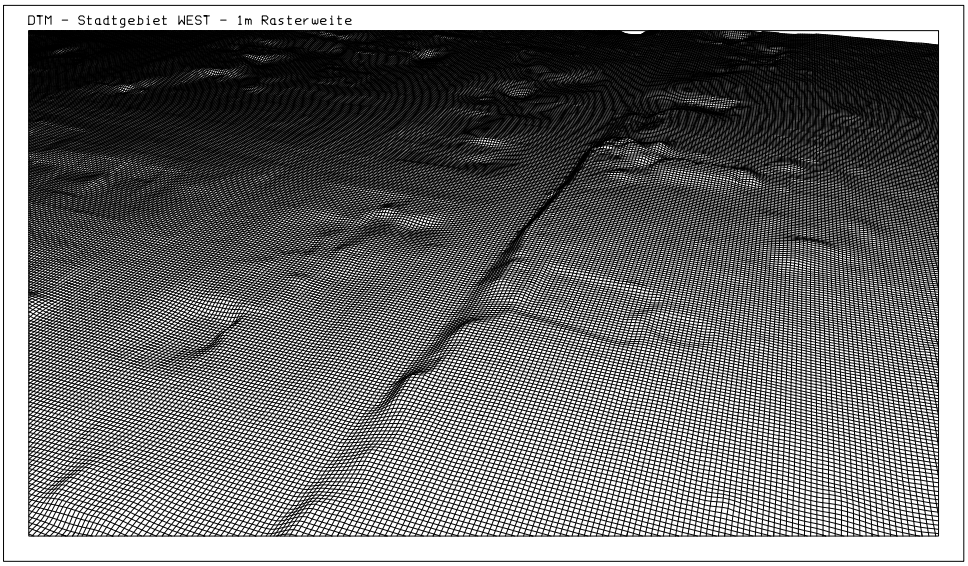


Figure 2.5: Final DTM (0.5 m; here: 1.0 m for better readability).

connected by small bridges are separated. The resulting binary image is analyzed by a connected component analysis which results in a greater number of regions, and the terrain roughness criterion is evaluated again. Pixels being in regions now classified as containing vegetation are erased in the initial building label image. Thus, in vegetation areas originally connected to buildings, only the border pixels remain classified as "building pixels". Again, morphological opening helps to erase these border pixels. The resulting building label image only contains a small percentage of erroneously classified pixels in some backyards (figure 2.6(e)).

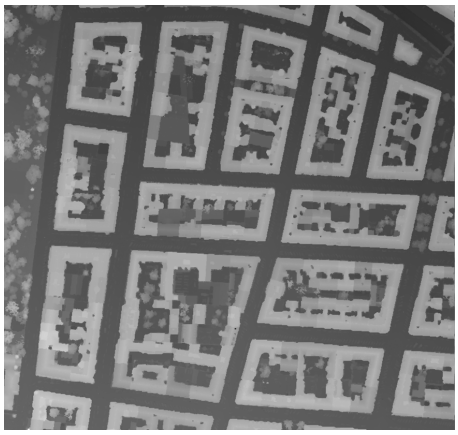
At a very coarse level of detail, a 3D city model can be derived by creating prismatic models from the boundary polygons of the building regions using the average building heights computed from the DSM. An example for such a city model with a height accuracy of about ± 5 m is shown in figure 2.6(f).

2.5 Geometrical Reconstruction of Buildings

2.5.1 Generation of Initial 3D Planar Segments

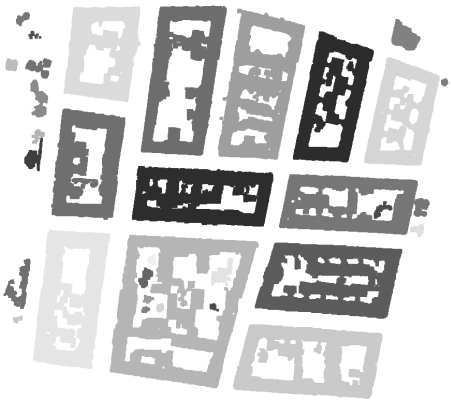
To start with model generation, initial 3D planar segments, their geometrical parameters, and their initial border polygons have to be found in the regions of interest. This is achieved by generating a "segment label image" defined in object space with an appropriate grid width. Each pixel of that image is assigned the label of the planar segment it belongs to.

The framework for polymorphic feature extraction [Fuchs, 1998] is applied for the generation of planar segments, too. Just as described in section 2.4, the framework is applied to the first derivatives of the DSM, this time using a small integration kernel of 3×3 pixels. Pixels classified as being homogeneous are surrounded by pixels having similar components of the normal vector, i.e., they are in a region containing co-planar points [Brunn and Weidner, 1997]. The binary image of the homogeneous pixels is used for further processing (figure 2.7(a)). By applying a connected component



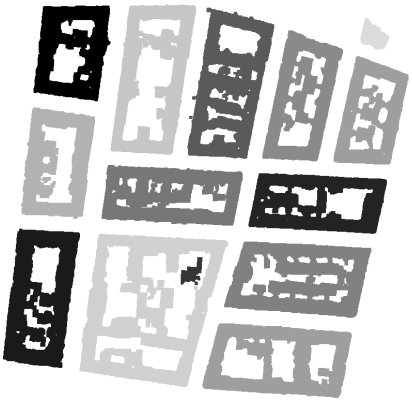
(a)

(b)



(c)

(d)



(e)

(f)

Figure 2.6: Building detection in a test site in the City of Vienna. Original resolution: 0.1 m (in-flight) by 1.0 m (cross-flight). (a) DSM; grid width: 0.5 $\times$ 0.5 m^2; extent: 410 $\times$ 435 m^2. (b) Initial building mask (height threshold h_{min} = 3.5 m). (c) Initial building label image before evaluating terrain roughness. (d) Results of texture classification. Integration kernel: 9 $\times$ 9 pixels. White: homogeneous; grey: linear; black: point-like. (e) Final building label image. Fourteen building regions have been detected. (f) VRML visualization of prismatic models created from the boundary polygons of the building regions from figure 2.6(e).

analysis to this binary image, planar patches should be detectable. However, due to classification errors, especially at the intersections of roof planes which are almost horizontal, the regions thus detected often turn out to be too large. Typically, this leads to L-shaped segments such as the one at the upper left corner of figure 2.7(a). In order to avoid these segmentation errors, an iterative strategy is applied for the generation of planar patches:

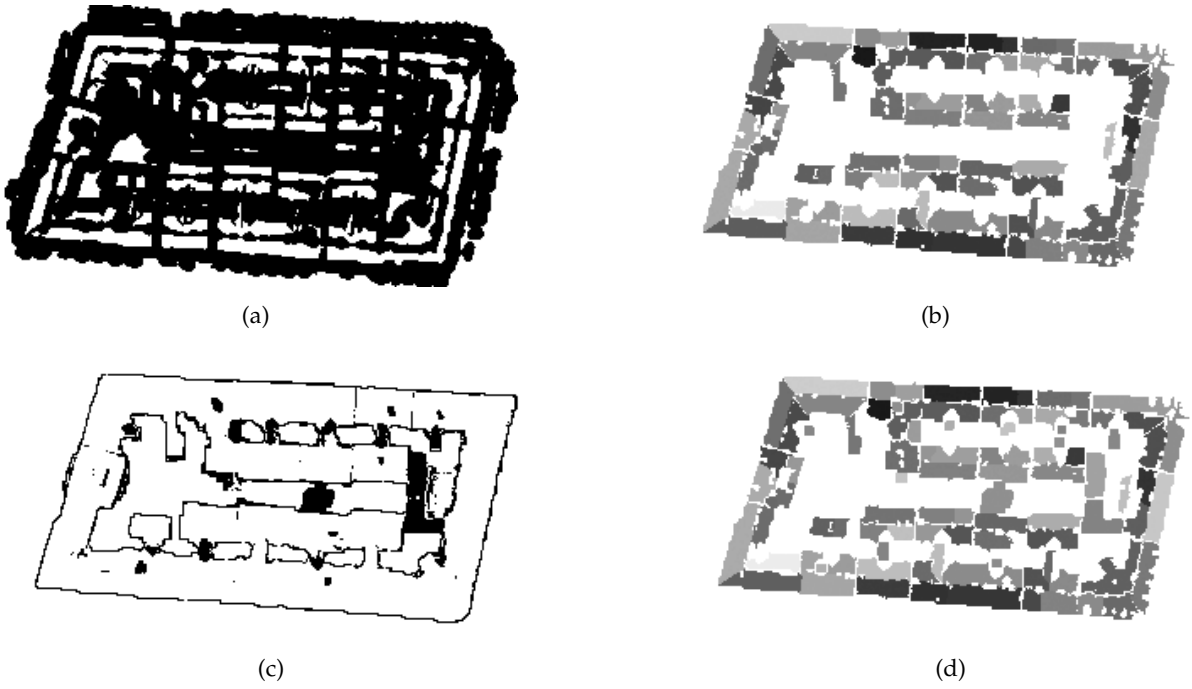
1. The binary image of homogeneous pixels is morphologically opened using a square structural element before applying the connected component analysis.
2. The geometric parameters of the planar patches thus detected are derived along with their r.m.s. errors from all points inside these patches.
3. The patches with the best fit, i.e., those with r.m.s. errors better than a certain threshold (e.g., $\pm$10 cm) are considered to be seed regions for region growing. These seed regions are grown iteratively by adding neighbouring pixels to a region if their distances from the original adjusting plane are below a certain threshold. In this way, the most relevant and best fitting planes are extracted from the DSM.
4. The plane parameters are updated, and the pixels already being assigned to a planar patch are erased in the binary image. The connected component analysis is repeated, and the parameters of the new planar patches are evaluated.

Steps 1 to 4 are repeated with a decreasing size of the structural element for morphological opening. Thus, smaller and smaller initial regions are found, and by only allowing well-fitting planes to grow, it is possible to split the regions corresponding to more than one roof plane, because the r.m.s. error of the planar fit is a good indicator for the occurrence of such situations. Figure 2.7(b) shows the planar patches extracted in one of the building regions from figure 2.6(e).

A further analysis has to detect planes which cover too small an area for resulting in pixels classified as being homogeneous. We search for regions not being consistent with the planar regions detected so far (figure 2.7(c)). The borders of the buildings are typically found in that process, which is caused by laser points on the walls and by the effects of grid interpolation. Again, we get rid of these points by a morphological opening operation using a 3×3 square structural element, and a connected component analysis is applied to the resulting image in order to create additional candidates for planar patches. Figure 2.7(d) shows the final segment label image created for one of the building regions from figure 2.6(e). The r.m.s. errors of planar adjustment varies between $\pm$5 cm and $\pm$15 cm for the segments corresponding to the "homogeneous" points. The segments having a r.m.s. error larger than $\pm$10 cm possibly still correspond to more than one roof plane. In the planar regions created by the analysis of the originally inconsistent points, the r.m.s. errors vary between $\pm$25 cm and $\pm$5 m. Some of these regions correspond to trees, and other regions still correspond to more than one roof plane. In the future, a further analysis will split these regions into smaller ones corresponding to even smaller planes in object space. This can be accomplished, e.g., by a height segmentation of the DSM in these regions. Table 2.1 shows the distribution of the r.m.s. errors of the planar fit.

2.5.2 Grouping Planar Segments to Create Polyhedral Models

To derive the neighbourhood relations of the planar segments, a Voronoi diagram based on a distance transformation of the segment label image has to be created [Ameri, 2000a]: each pixel inside the re-



(a) (b) (c) (d)

Figure 2.7: (a) Pixels classified as "homogeneous" (white) for one of the building regions in figure 2.6(e). (b) Planar patches obtained by iteratively applying region growing. (c) Pixels not being consistent to a planar patch. (d) Final segment label image for one of the building regions in figure 2.6(e).

r.m.s. error [m]	Regions	Pixels [%]
0.00 - 0.05	241	30.6
0.05 - 0.10	333	44.4
0.10 - 0.15	96	9.3
0.15 - 0.20	133	8.8
0.20 - 0.50	26	0.9
0.50 - 1.00	10	0.2
1.00 - 2.00	14	0.5
2.00 - 3.00	42	2.6
3.00 - 4.00	36	1.8
>= 4.00	15	0.9

Table 2.1: Distribution of the r.m.s. errors of the planar fit. Regions: number of planar regions in the respective range of r.m.s. errors. Pixels: percentage of pixels in these regions compared to the number of all pixels in all planar regions. 68% of the pixels in the building candidate regions are classified as belonging to a planar region.

gion of interest not yet assigned to a planar segment is assigned to the nearest segment. The distances of pixels from the nearest segment are computed by using a 3-4 chamfer mask. Figure 2.8 shows a Voronoi diagram of the segment label image from figure 2.7(d). From the Voronoi diagram, the neighbourhood relations of the planar segments are derived, and the borders of the Voronoi regions can be extracted as the first estimates for the border polygons of the planar segments (figure 2.9).

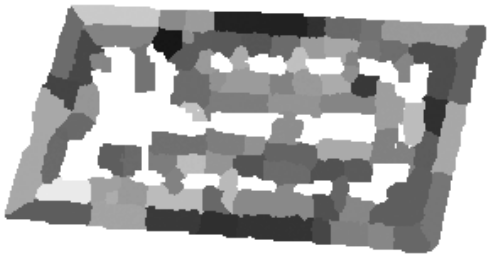


Figure 2.8: A Voronoi diagram of the label image in figure 2.7(d).

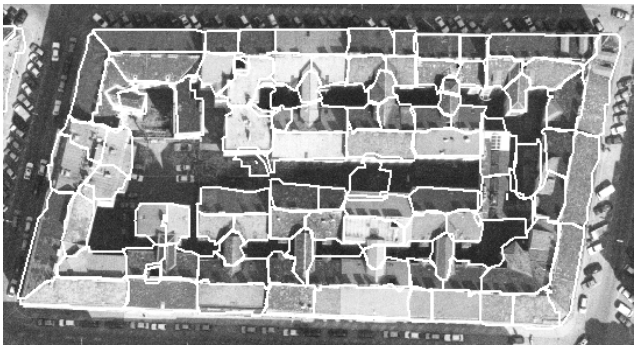


Figure 2.9: The roof polygons of the building in figures 2.7 and 2.8 back-projected to an aerial image.

After deriving the neighbourhood relations, neighbouring planar segments have to be grouped. There are three possibilities for the relations of two neighbouring planes [Baillard et al., 1999]. First, they might be co-planar, which is found out by a statistical test applied to the plane parameters. In this case, they have to be merged. Second, two neighbouring planes might intersect consistently, which is the case if the intersection line is close to the initial boundary. In this case, the intersection line has to be computed, and both region boundaries have to be updated to contain the intersection line. Third, if the planes do not intersect in a consistent way, there is a step edge, and a vertical wall has to be inserted at the border of these segments. After grouping neighbouring planes, the bounding polygons of all enhanced planar regions have to be completed. [Moons et al., 1998] give a method for doing so and for regularizing the shape of these polygons at building corners. Finally, the planar polygons have to be combined to form a polyhedral model, and vertical walls as well as a floor have to be added to the model.

The tools for grouping planes and for computing intersections and the positions of step edges have not yet been implemented. Figure 2.10 shows a VRML visualization of a 3D model created from intersecting vertical prisms bounded by the borders of the Voronoi regions with the respective 3D roof planes. The structure of the roofs is correctly resembled, but the intersection lines of neighbouring roof planes are not yet computed correctly. However, the visualization shows the high potential of the method for generating roof planes from LIDAR data.

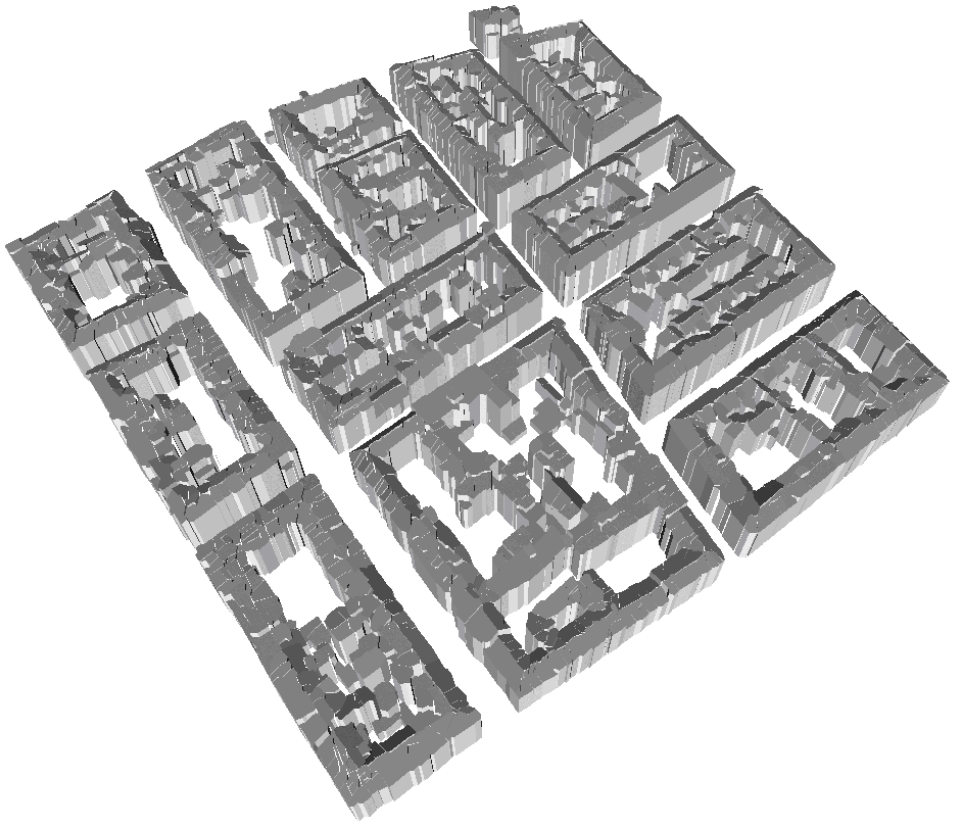


Figure 2.10: VRML visualization of a model created from the boundary polygons of the Voronoi diagrams of the roof planes extracted for all building regions in figure 2.6(e).

2.6 Conclusion and Future Work

We have presented a method for the extraction of buildings from high-resolution LIDAR data in densely built-up areas. Preliminary results from a test site in the City of Vienna have shown the high potential of the method. Future work will not only include the implementation of the modules still missing, but also the assessment of quality parameters for the results. With respect to the building outlines, this can be accomplished by a comparison to existing GIS data. In addition, we will perform research work in order to improve the results of building extraction by additionally using digital aerial images for matching of roof edges with image edges.

Acknowledgements

This research has been supported by the Austrian Science Foundation (FWF) under project no. P14083-MAT.

Chapter 3

Building Detection Using LIDAR Data and Multi-spectral Images

3.1 Introduction

3.1.1 Motivation and Goals

Automation in data acquisition for 3D city models is an important topic of research in photogrammetry. In addition to photogrammetric techniques relying on aerial images, the generation of 3D building models from point clouds provided by airborne laser scanning, also known as LIDAR (*LI*ght *D*etection *A*nd *R*anging), is gaining importance. This development has been triggered by the progress in sensor technology which has rendered possible the acquisition of very dense point clouds using airborne laser scanners [Kraus, 2002]. Using high-resolution LIDAR data, it is not only possible to detect buildings and their approximate outlines, but also to extract planar roof faces and, thus, to create models correctly resembling the roof structure [Brenner, 2000, Rottensteiner and Briese, 2002, Vosselman and Dijkman, 2001].

With decreasing resolution, it becomes more difficult to correctly detect buildings in LIDAR data, especially in residential areas characterised by detached houses. In order to improve the performance of building detection, additional data can be considered:

- LIDAR systems register two echoes of the laser beam, the *first* and the *last pulse*. If the laser beam is reflected at the bare soil, first and last pulse will refer to the same object point. If the laser beam hits a tree, a part of the light will be reflected at the canopy, resulting in the first pulse registered by the sensor. The rest will penetrate the canopy and, thus, be reflected further below, maybe even at the soil. The last pulse registered by the sensor corresponds to the lowest point where the signal was reflected [Kraus, 2002].

- Along with the run-time of the signal, the *intensity* of the returned laser beam is registered by LIDAR systems. LIDAR systems typically operate in the infrared part of the electromagnetic spectrum. Unfortunately, given the footprint size of the laser beam (e.g. 2-3 dm) and the average point distance (e.g. 1-2 m), the intensity image is undersampled and, thus, very noisy [Vosselman, 2002].

- As building detection is a classification task, *multi-spectral images* can provide valuable information due to their spectral content and because their resolution is still better than the resolution of laser scanner data [Schenk and Csatho, 2002].

Apart from the problems related to sensor resolution, building detection is also made complicated by the fact that buildings may have quite different appearances both with respect to their geometric dimensions and their reflectance properties. Thus, it is often impossible to select appropriate thresholds and filter sizes for various algorithms. It is a well-known strategy in image matching to apply algorithms to data having a lower resolution first to get approximate values, refining these results in an iterative way, in each iteration considering data of a better resolution than in the previous one, until the original resolution is reached. It is the goal of this chapter to present a new method for the automatic detection of buildings of heterogeneous appearance from LIDAR data and multi-spectral images making use of such a hierarchic approach. With respect to the combination, or fusion [Klein, 1999] of these data, we currently apply heuristic rules in the detection process. In the future, the proposed algorithm is to be used as a module giving building candidate regions for a framework for feature based data fusion as it is described in [Lu and Trinder, 2003]. The examples presented in this chapter were computed using the LIDAR and image data from a test site in the Fairfield (New South Wales) covering an area of 2×2 km^2.

3.1.2 Related Work

There have been several attempts to detect buildings in LIDAR data in the past. The task has been solved by classifying the LIDAR points according to whether they belong to the terrain, to buildings or to other object classes, e.g., vegetation. Morphological opening filters or rank filters are used to determine a digital terrain model (DTM) which is subtracted from the digital surface model (DSM). By applying height thresholds to the normalized DSM thus created, an initial building mask is obtained [Weidner, 1997]. The initial classification has to be improved in order to remove vegetation areas. In [Brunn and Weidner, 1997], this is accomplished by a framework for combining various cues in a Bayesian network. They also use a hierarchic strategy, turning the classification results of the coarser resolution into one of the cues for the classification in the next iteration. The problem with their approach is related the complexity of estimating the conditional probabilities required for the Bayesian network. In [Weidner, 1997], Weidner has tackled he problem of precisely determining the building outlines by applying the minimum description length principle for deciding on regularizations.

In [Lu and Trinder, 2003], a DSM derived by image matching and a colour image are fused on the basis of the Dempster-Shafer theory. Fusion is carried out at feature level, the initial segmentation being performed by a K-means unsupervised classification of the colour images, using cues such as the normalised difference vegetation index (NDVI) and the average relative height of the feature to distinguish buildings from other objects. In order to overcome the deficiencies of the method used to obtain the initial segmentation, it would be desirable to have the DSM take over a more prominent role in the process, which is even more advisable if the DSM is not created by image matching (which involves some sort of smoothing), but derived from the LIDAR data.

In [Rottensteiner and Briese, 2002], we have presented an algorithm for building detection that relied on DTM generation by hierarchic robust linear prediction [Briese et al., 2002], using the DTM and DSM grids for further classification. The method has been shown to give good results in densely built-up areas [Rottensteiner and Briese, 2002], but in more heterogeneous areas containing houses of different sizes and also forests, tuning the parameters for DTM generation is not an easy task. In

this chapter, we will describe how that method has been modified to consider the additional data sources and to work in a hierarchic way without relying on hierarchic robust linear prediction.

3.2 Work Flow for Building Detection

The work flow for our method is depicted in figure 3.1. The input to our method is given by three data sets that have to be generated from the raw data in a pre-processing step. The *last pulse DSM* is sampled into a regular grid by linear prediction using a straight line as the covariance function, so that the interpolation is carried out almost without filtering [Rottensteiner and Briese, 2002]. The *first pulse DSM* is also sampled into a regular grid. The *normalised difference vegetation index (NDVI)* is computed from the near infrared and the red bands of the multi-spectral images we assume to be available [Lu and Trinder, 2003]. These image data must be geocoded.

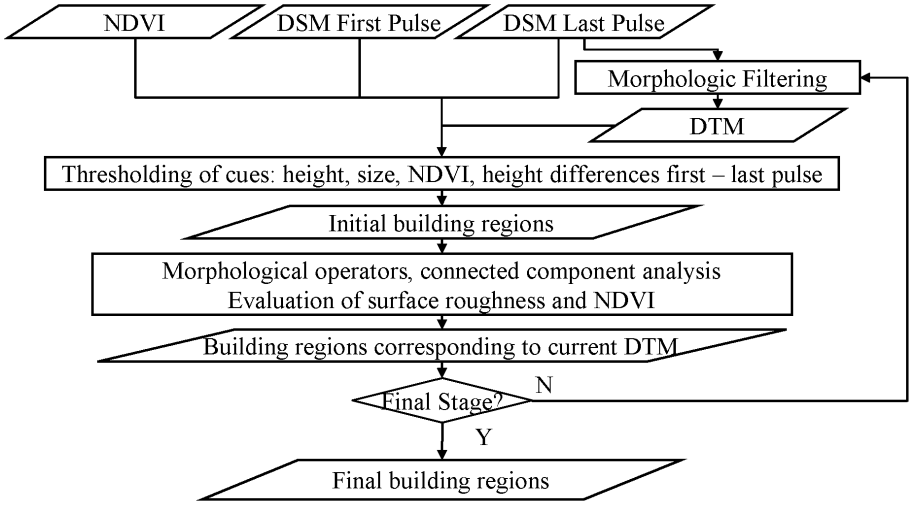


Figure 3.1: The work flow for building detection.

We start by creating a DTM from the last pulse DSM by morphological grey scale opening using a square structural element. Initially, the size of the structural element corresponds to the size of the largest building available in the dataset. An initial building mask, basically a binary image of possible building pixels, is created mainly by thresholding operations. That building mask is morphologically opened to eliminate spurious and strangely shaped building areas, and then connected components of building pixels give the initial building regions. For these regions, we evaluate the surface roughness and the average NDVI. In the first iterations, very tight thresholds are applied to surface roughness, because we assume the largest buildings in the scene to consist of large planar surfaces. Thus, we obtain an intermediate set of building regions, only containing the largest and most salient buildings (corresponding to the current state of the DTM). After that, the whole procedure is repeated with a smaller structural element for morphological opening, but in the areas already classified as buildings, the DTM heights of the previous iteration are substituted for the results of the morphological filter. Thus, the buildings already detected are eliminated, whereas the smaller size of structural element for morphological filtering helps to obtain a finer approximation for the DTM, rendering possible the separation of smaller buildings. The whole procedure is repeated with

a decreasing size of the structural element until a user-specified minimum size is reached. The results of the final iteration are identical to the results of building detection, basically represented by a list of building regions and a building label image. In section 3.3, we will have a closer look at the individual processing stages.

3.3 Stages of Building Detection

3.3.1 Morphological Filtering of the DSM

We assume the DSM to be a matrix containing the heights $z(i, j)$, with integer matrix indices i and j. For morphological filtering of the DSM, a structural element W, i.e., a digital image $w(m, n)$ representing a certain shape, has to be defined. Restricting ourselves to symmetrical structural elements having constant values, thus $w(m, n) = w(-m, -n)$ and $w(m, n) = 0\ \forall [m, n] \in W$, morphological opening is performed by first carrying out an morphological erosion,

$$\overline{z}(i, j) = \min_{[m,n] \in W} z(i - m, j - n) \tag{3.1}$$

followed by a morphological dilation,

$$\overline{\overline{z}}(i, j) = \max_{[m,n] \in W} \overline{z}(i - m, j - n) \tag{3.2}$$

yielding the morphologically opened height matrix $\overline{\overline{z}}(i, j)$ [Weidner, 1997]. The resulting height matrix does not contain objects smaller than the structural element W. If W is chosen to be greater than the largest building in the data set, all buildings are removed by morphological opening; however, if W is too large, terrain details might be removed, too. If W is chosen rather small, the results of morphological filtering will be closer to the original height matrix and, thus, to the terrain, but larger buildings will remain in the data set. This is the reason why we apply morphological filtering in the hierarchical framework described in section 3.2.

3.3.2 Generation of the Initial Building Label Image

Morphological filtering provides us with an approximation for the DTM. As described in section 3.2, for buildings already detected, the DTM generated by morphological opening in the previous iteration is substituted for the results of morphological filtering, so that large buildings that would be preserved by morphological filtering in the current iteration are eliminated beforehand. An initial building mask is created by thresholding the height differences between the last pulse DSM and the DTM (e.g., by h_{min}=2.5 m). This initial building mask still contains spurious regions, areas covered by vegetation, and terrain structures smaller than the structural element for morphological filtering. In addition, individual buildings might not be separated correctly.

At this instance, the additional data sources can be used to improve these results. First, a large NDVI indicates areas covered by vegetation, so that pixels having an NDVI above a certain threshold are erased in the building mask. Second, as pointed out in section 3.1.1, the heights from first and last pulse data will differ mainly in areas covered by trees and if the laser beam accidentally hits the roof edge of a building. Thus, in most cases, large height differences between first and last pulse

data indicate trees. Pixels having a height difference larger than a certain threshold are erased in the building mask, too.

A binary morphological opening filter using a structural element of a size corresponding to the expected minimum size of a building part (e.g., 3×3 m^2) is applied to the initial building mask to erase oddly shaped objects such as fences and to separate building regions just bridged by a thin line of pixels. A connected component analysis of the resulting image is applied to obtain the initial building regions. Regions smaller than a minimum area are discarded.

3.3.3 Classification of Building Candidate Regions

Some of the initial building regions correspond to groups of trees or to terrain structures smaller than the structural element. These regions can be eliminated by evaluating a surface roughness criterion derived by an analysis of the second derivatives of the DSM. In [Fuchs, 1998], a method for polymorphic feature extraction is described which aims at a classification of texture as being homogeneous, linear, or point-like, by an analysis of the first derivatives of a digital image. The thresholds required for that classification are derived automatically from the image data. This method is applied to the first derivatives of the DSM using an integration kernel of a size corresponding to, e.g., 5 m in object space. Under these circumstances, "homogeneous" pixels correspond to areas of locally parallel surface normal vectors, thus, they are situated in a locally planar neighbourhood. "Linear" pixels correspond to the intersections of planes. Finally, "point-like" pixels are in a neighbourhood of great, but anisotropic variations of the surface normal vectors. This is typical for building corners and for trees. For evaluating surface roughness, the numbers of "homogeneous" and "point-like" pixels are counted in each initial building region. Buildings are characterised by a large percentage of "homogeneous" and by a small percentage of "point-like" pixels. By comparing these percentages to thresholds, non-building regions can be eliminated. The surface roughness criterion works well for large buildings and with dense LIDAR data [Rottensteiner and Briese, 2002]. If the point distance of the LIDAR data is larger than, e.g., 1 m, only few LIDAR points will be situated on small buildings, so that the percentage of "homogeneous" pixels is reduced, whereas the percentage of "point-like" pixels is increased. Thus, it makes sense to additionally evaluate the average NDVI for each building region to eliminate vegetation areas.

Finally, vegetation areas still connected to buildings are eliminated. By morphological opening, regions just connected by small bridges are separated. The resulting binary image is analysed by a connected component analysis which results in a greater number of regions, and the terrain roughness criterion is evaluated again. Pixels being in regions now classified as containing vegetation are erased in the initial building label image. Thus, in vegetation areas originally connected to buildings, only the border pixels remain classified as "building pixels". Again, morphological opening helps to erase these border pixels [Rottensteiner and Briese, 2002].

3.4 Experiments

3.4.1 Description of the Data Set

The test data set was captured over Fairfield (New South Wales) using an Optech ALTM 3025 laser scanner. Both first and last pulses were recorded, the average point distance being about 1.2 m. We derived DSM grids at a resolution of 1 m from these data. Intensity data were available, too.

We used an area of 2 $\times$ 2 km^2 for our test. For that area, a true colour digital orthophoto with a resolution of 15 cm was available. The orthophoto was created using a DTM, so that the roofs and the tree-tops were slightly displaced with respect to the LIDAR data. Unfortunately, the digital orthophoto did not contain an infrared band. We circumvented this problem by resampling both the

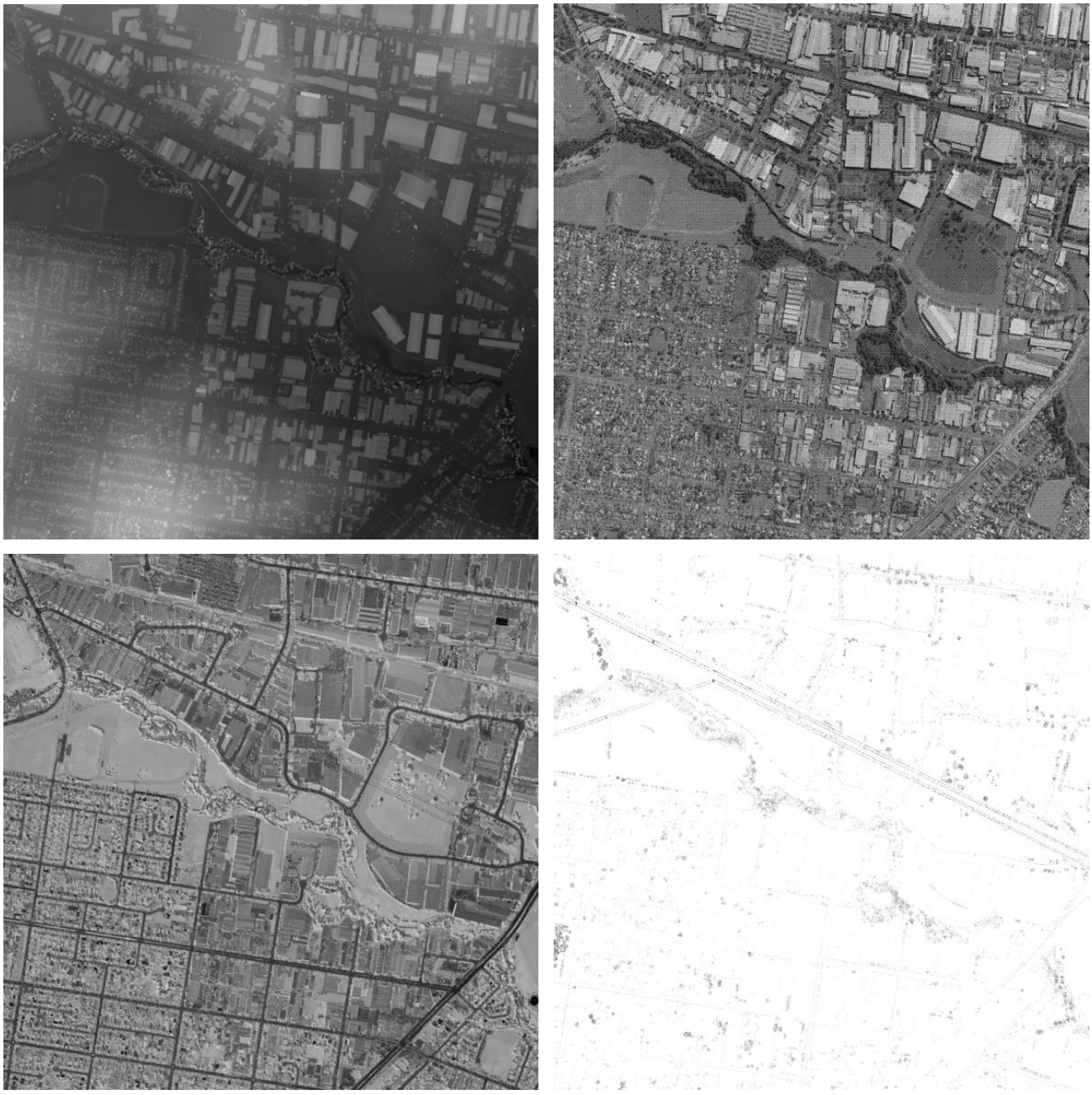


Figure 3.2: The Fairfield data set. Upper row, left: The DSM (black: low areas, white: high areas). Upper row, right: Intensity of the colour orthophoto. There are industrial buildings in the northern and central regions and residential houses in the west and south-west. Second row, left: The "pseudo-NDVI-image". Second row, right: The height differences between first and last pulse data (black: large differences). Total area: 2000 $\times$ 2000 m^2.

digital orthophoto and the (infrared) LIDAR intensity data to a resolution of 0.5 m and by computing a "pseudo-NDVI-image" from the LIDAR intensities and the red band of the digital orthophoto. Apart from problems with georeferencing caused by the displaced tree canopies in the orthophoto, there were also problems with shadow areas in the orthophoto, so that the "pseudo-NDVI-image" did not provide as valuable an information as the NDVI image was supposed to be. Still, it helped in classification. The input data for our test are shown in figure 3.2.

3.4.2 Results

Figure 3.3 shows the results of morphological opening of the DSM in figure 3.2 using structural elements of two different sizes (150 m and 25 m). Using the large structural element, all buildings can be eliminated, but the terrain shape is not well preserved, so that the residential buildings in the lower part of the scene are merged. This can be seen in figure 3.4, showing the initial building mask derived using the left DTM in figure 3.3 (height threshold: 2 m). Using a smaller structural element, more terrain details are preserved, but the large buildings are still contained in the data. However, using our hierarchic approach, these buildings can be eliminated beforehand, as described in section 3.2.

The results of texture classification are presented in figure 3.5 (filter kernel size: 3×3 pixels). Note the co-incidence of clusters of "point-like" pixels and vegetation areas such as those along the creek passing through the scene diagonally, which is used to eliminate trees. In the first iteration, starting from the initial building mask in figure 3.4, we only accept regions larger than 2500 m^2, containing less than 30% of "point-like" and at least 70% of "homogeneous" pixels, thus, large regions consisting of mostly planar roof planes. As the industrial buildings had a high reflectivity in the infrared part of the spectrum, the threshold for the "pseudo-NDVI" was kept at 75%. 85 mostly large

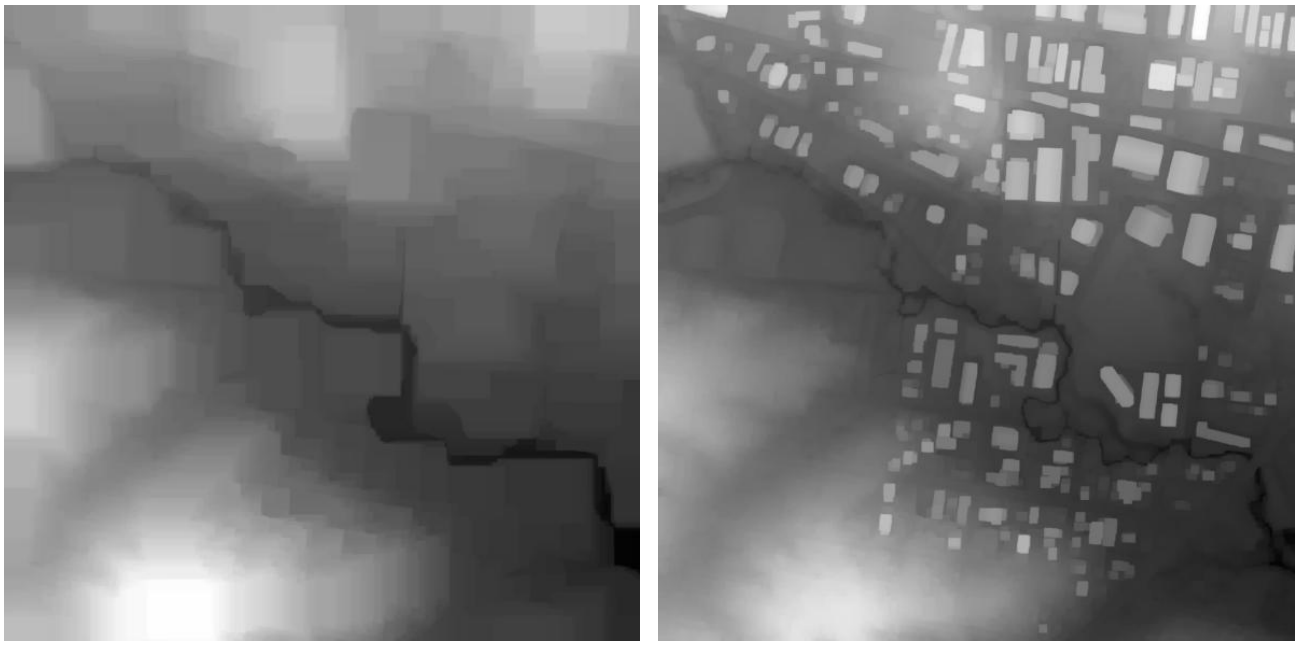


Figure 3.3: The results of morphological opening of the DSM using structural elements of 150×150 m^2 (left) and 25×25 m^2 (right).

Figure 3.4: The initial building mask corresponding to the left DTM in figure 3.3. The residential buildings on a small hill at the southern border are merged.

Figure 3.5: The classification results of polymorphic feature extraction. White: "homogeneous", light grey: "linear", black: "point-like".

building structures are detected in the first iteration (left part of figure 3.6).

Altogether three iterations were carried out, using structural elements of 150 m, 75 m, and 25 m. In the final iteration, regions larger than 25 m^2, containing less than 85% of "point-like" and at least 1% of "homogeneous" pixels were accepted. These relatively loose settings of the threshold were a consequence of the LIDAR resolution, with only few points and, thus, few "homogeneous" pixels on the roofs of the smaller buildings. 1589 buildings were detected (right part of figure 3.6). The parameters for classification were chosen in a way to minimise the number of missed buildings, accepting a larger rate of false alarms. Less than 1% of the buildings were not detected. [1]

Figure 3.7 shows the orthophoto of a part of the test area super-imposed by the boundary polygons of the buildings. There remain some trees in the data, and some of the buildings are still merged, especially if the distance between them is small. However, as almost all the buildings are contained in the data, it might be possible to improve the results of classification by considering additional cues derived, for instance, from the colour images.

3.5 Conclusions and Future Work

We have presented a hierarchic method for building detection from LIDAR data and multi-spectral images, and we have shown its applicability in a test site of heterogeneous building shapes. In our test, we put more emphasis on detecting all buildings in the test data set than on reducing the

[1] This number does not consider small building structures such as garden sheds and is thus not comparable to the completeness rates presented in chapters 4 and 5.

Figure 3.6: Buildings extracted in the first (left) and third (right) iterations.

Figure 3.7: Boundary polygons super-imposed to the orthophoto. A colour version can be found on page 129.

false alarm rate because in the future we want this method to be the module for initial segmentation in a framework using more sophisticated methods of data fusion similar to those described in [Lu and Trinder, 2003], replacing the rather heuristic methods used up to now. In order to further improve the segmentation results and to split building regions erroneously merged, the results of a segmentation of the orthophoto could be considered. Using the NDVI computed from real infrared images rather than the "pseudo-NDVI" used in this test might also help to improve the results.

Acknowledgements

This work was supported by the Australian Research Council (ARC) under Discovery Project DP0344678. The Fairfield data set was provided by AAM Geoscan, Kangaroo Point, QLD 4169, Australia (www.aamgeoscan.com.au).

Chapter 4

Using the Dempster-Shafer Method for the Fusion of LIDAR Data and Multi-spectral Images for Building Detection

4.1 Introduction

4.1.1 Motivation and Goals

Automation in data acquisition for 3D city models is an important topic of research in photogrammetry. In addition to aerial images, point clouds generated by airborne laser scanning, also known as LIDAR (LIght Detection And Ranging), are being used for that purpose more frequently. This development has been triggered by the progress in sensor technology, which has rendered possible the acquisition of very dense point clouds using airborne laser scanners [Kraus, 2002]. Problems to be tackled in this context comprise the generation of high quality digital terrain models (DTMs) in urban areas and the extraction of topographic objects such as buildings or trees. These problems are closely interrelated. For computing a DTM, the LIDAR points on the tops of buildings and trees have to be eliminated, and thus information about the positions of such objects is required, whereas on the other hand, a DTM is required if buildings or trees are to be detected. This means that the first step to be carried out in order to extract meaningful semantic information from LIDAR data is a classification of data to separate points situated on the terrain from those on buildings, trees, and other objects. With high-resolution LIDAR data, it is not only possible to perform this classification, but also to reconstruct objects such as buildings geometrically at a relatively high level of detail [Haala and Brenner, 1999, Rottensteiner and Briese, 2002, Vosselman and Dijkman, 2001].

With the decreasing resolution of the LIDAR data, the classification becomes more difficult, especially in residential areas characterised by detached houses, where the appearance of trees and buildings in the data might be similar. In order to improve the classification results, additional data sources can be considered. First, LIDAR systems register two echoes of the laser beam, the *first* and the *last pulse*, corresponding to the first and the last points from where the laser beam is reflected. Differences between the first and last pulse data indicate height steps, e.g. at building boundaries,

power lines, and, most frequently, trees [Kraus, 2002]. Second, the *intensity* of the returned laser beam can be used. LIDAR systems typically operate in the infrared part of the electromagnetic spectrum. However, these data have the disadvantage that they are under-sampled and, thus, very noisy [Vosselman, 2002], because the footprint size of the laser beam is typically 0.2 m, while the average point distance is 1 m. Third, *multi-spectral images* can provide valuable information due to their spectral content and because their resolution is better than the resolution of laser scanner data [Schenk and Csatho, 2002].

It is the goal of this chapter to investigate the fusion of first and last pulse LIDAR data and multi-spectral images for building detection in densely built-up urban areas. In the course of the building detection process, fusion will be especially helpful for improving the classification of land cover. Although our emphasis is on the detection of buildings, we also want to distinguish three other classes that are relevant in the context of 3D city models and for the generation of high-quality DTMs in urban areas: trees, grassland, and bare soil. This is accomplished by first applying a hierarchical technique for coarse DTM generation, followed by the fusion of various cues derived from the input data for classification. Here, data fusion is based on the theory of Dempster-Shafer [Klein, 1999]. Examples are presented for a test site in Fairfield (New South Wales) covering an area of 2×2 km^2.

4.1.2 Background

Existing classification techniques for building detection from LIDAR and/or multi-spectral data can be characterised by the cues that are actually used for classification and by the methods used for the data fusion. These topics will be discussed separately.

4.1.2.1 Cues for Building Detection

From the LIDAR data, a *digital surface model* (DSM) can be derived, for example by sampling the data into a regular grid [Rottensteiner and Briese, 2002, Maas, 1999]. The DSM represents the surface from which the laser pulse is reflected, that is trees, terrain surface, buildings, etc. In flat terrain, the elevations of the DSM can be used directly to separate elevated objects from others. In the case of undulating terrain, the elevations reflect both the terrain heights and the height differences between points on elevated objects and the terrain. That is why the classification of LIDAR data for the automatic detection of topographic objects usually starts with the generation of a DTM, e.g. by morphological opening filters or rank filters [Weidner and Förstner, 1995] or by hierarchical robust linear prediction [Briese et al., 2002]. We apply a hierarchical method of morphological filtering for DTM generation in this study [Rottensteiner et al., 2003], which will be outlined in section 4.2.3. The DTM is subtracted from the DSM, yielding a *'normalised' DSM*, the heights of which directly reflect the heights of objects relative to the terrain [Weidner and Förstner, 1995].

A DSM also provides information about *local surface properties* via an analysis of the derivatives of the DSM. The *maximum slope* has been used for distinguishing flat roofs from tilted roofs [Maas, 1999]. The second derivatives are more commonly used for building detection. As they are closely related to the local curvature of the DSM, they can be used to derive measures for *surface roughness*. Assuming that roofs mostly consist of planar or at least smooth surfaces, an analysis of surface roughness can help to separate buildings from trees. Various parameters have been used in the past to characterise surface roughness, e.g. the output of a Laplace filter applied to the DSM [Maas, 1999, Vögtle and Steinle, 2003], local curvature [Vögtle and Steinle, 2003], or the local variance of the surface normal vectors [Brunn and Weidner, 1997]. The model for surface roughness we

use is based on an analysis of the local variations of the surface normal vectors using the framework of polymorphic feature extraction [Förstner, 1994]. It will be described in section 4.2.2.

Height differences between first and last pulse data have also been used to improve the results of building detection. They can be used as an indicator of vegetation, but large differences also occur at the edges of buildings. Until recently, laser scanners only could deliver either first or last pulse data, and the user of the data had to select one of these scanning modes according to his or her priorities. Modern laser scanners can deliver both first and last pulse data in the same flight. A detailed discussion of the effects of choosing one of the two scanning modes on the results of building detection is given in [Steinle and Vögtle, 2000]. In addition to the local characteristics described up to now, *average surface roughness parameters* or the *size of building candidate regions* can be evaluated to eliminate candidate regions being either oddly shaped, too large, too small, or characterised by a large average surface roughness [Rottensteiner and Briese, 2002, Weidner and Förstner, 1995].

Schenk and Csatho [Schenk and Csatho, 2002] put forward the idea of exploiting the complementary properties of LIDAR data and *aerial imagery* to first achieve a more complete surface description by a feature based fusion process and then to extract semantically meaningful information from the aggregated data. They pointed out that LIDAR data are especially useful for the detection of surface patches having specific geometrical properties, and for deriving other properties such as their roughness. On the other hand, aerial images can help to provide the surface boundaries and the locations of surface discontinuities.

Haala and Brenner [Haala and Brenner, 1999] combine a normalised DSM from LIDAR data with the three spectral bands of a scanned *colour infrared (CIR) image*. As the separation of trees from buildings is the most problematic task in this context due to the relatively low resolution of the LIDAR data, it is important to include the near-infrared band in the classification process. As an alternative to using all bands of multi-spectral images, the *normalised difference vegetation index (NDVI)* derived from the near infrared and the red portions of the spectrum can be applied for its potential in discriminating vegetation [Lu and Trinder, 2003].

4.1.2.2 Data Fusion Techniques for Building Detection

Building detection is usually carried out in two stages:

1. *Detection of building candidate segments:* An initial classification is carried out on a per-pixel level. Each pixel is classified according to whether it is a building candidate pixel or not. The simplest way of doing so is applying a height threshold to the normalised DSM [Weidner and Förstner, 1995], but more sophisticated classification techniques can be applied, too. Connected components of building pixels then become building candidate regions.

2. *Evaluation of the building candidate regions:* For the initial regions, average parameters describing the DSM heights, the spectral characteristics, surface roughness, and the region size are also evaluated to separate the buildings from the trees [Rottensteiner and Briese, 2002, Weidner and Förstner, 1995, Lu and Trinder, 2003]. Again, various techniques can be applied to combine the available cues.

The results of an object-based classification, which is applied at the second stage above, cannot be better than the results of the initial segmentation. For instance, if trees standing close to buildings are merged in the initial segmentation, they can no longer be separated. In

[Rottensteiner and Briese, 2002], we have shown how the segmentation results can be improved at a later processing stage, but in general it would be preferable to improve the initial segmentation results by applying better fusion techniques in the first stage of classification. Some post-processing will also be required in that case: per-pixel classifiers do not consider the local context of a pixel or relationships with neighbouring pixels, thus there will be isolated spurious pixels. The classification results are especially unreliable at the building boundaries [Maas, 1999].

The simplest way to combine the different cues at each pixel is to concatenate the data from each cue to form a multi-dimensional data vector. The resultant vector can be treated as if it were from a single source [Lee et al., 1987]. That data vector can be used in standard procedures such as supervised maximum likelihood (ML) classification, unsupervised classification, or rule-based classification. As an alternative, Le Hégarat-Mascle et al. [Le Hégarat-Mascle et al., 1997] name three more sophisticated techniques: fuzzy logic, probabilistic reasoning, and the theory of Dempster-Shafer. In the remainder of this section we want to give an overview of how these techniques have previously been used for building detection.

Supervised ML classification is applied to the DSM heights, slopes, and the results of a Laplace filter by Maas [Maas, 1999]. He uses sparse training regions containing the object classes 'flat roof', 'tilted roof', 'vegetation', 'flat terrain', and 'no data', but states that the selection of training regions might be replaced by introducing a priori knowledge about the class centres and covariance matrices to replace the interactive selection of training regions. However, it has already been pointed out in [Lee et al., 1987] that this may not be suitable when the various sources cannot be described by a common "spectral" model, especially when spectral and elevation data are combined. For instance, neither the relative heights nor the spectral characteristics of buildings can be assumed to be normally distributed. Buildings have different heights and spectral "colours", so that they correspond to more than one cluster in feature space.

Unsupervised classification based on the ISODATA algorithm is applied to the three bands of a CIR image and a normalised DSM by Haala and Brenner [Haala and Brenner, 1999]. They stress the importance of the elevation data for the separability of the feature classes. The interpretation of the detected feature clusters is performed interactively. Appropriate modelling of the relevant classes to perform that interpretation automatically seems to be difficult. Lu and Trinder [Lu and Trinder, 2003] applied a K-means algorithm to RGB images to obtain an initial segmentation and evaluate additional sources such as the NDVI and a DSM derived from image matching to automatically assign the feature clusters to thematic classes. The main problem with their approach is that the DSM is not used in the original unsupervised classification, so that shadows cause relevant building parts to be missed entirely.

Rule-based classification is based on expert knowledge about the appearance of certain object classes in the data that are used to define rules by which the classes can be separated. These rules often involve thresholding operations, as in the method we will outline in section 4.2.3. Selecting "hard" thresholds properly is a critical issue. In addition, the hierarchical structure of many rule-based approaches, where first a subset of the cues is selected to make an initial classification and then the other cues are used to resolve any ambiguities [Lee et al., 1987], makes it impossible to recover from previous errors in the classification process. For this reason, we believe that algorithms evaluating all cues simultaneously are preferred, and even more so if the reliability of the results can be quantified.

Fuzzy logic can be used to model vague knowledge about class assignment in order to avoid hard thresholds as in rule-based algorithms [Binaghi et al., 1997]. In [Vögtle and Steinle, 2003], the

membership functions for a fuzzy logic classification using various cues derived from first and last pulse LIDAR data are described. These membership functions had to be defined for every class in a training phase. In a second step, these membership values were combined to obtain a final decision. This fuzzy classification is applied to previously detected building candidate segments, rather than on a per-pixel basis, and the method was used in a supervised context.

Probabilistic reasoning aims at assigning each pixel to the class $C_i \in \{C_1, \ldots, C_N\}$ maximising the a posteriori conditional probability $P(C_i|\mathbf{x}_s)$ of C_i given the data vector $\mathbf{x}_s$ for a pixel or feature s. The conditional probabilities are computed using the theorem of Bayes [Gorte, 1999] which in turn requires the modelling of the a priori probabilities $P(\mathbf{x}_s|C_i)$ of data vector $\mathbf{x}_s$ under the assumption of a class C_i and $P(C_i)$ of class C_i. The probabilities $P(\mathbf{x}_s|C_i)$ are often modelled by a multivariate Gaussian distribution. Initially, the prior $P(C_i)$ is often assumed to be equal for all classes, and then iteratively recomputed from the relative numbers of pixels in the first iteration. As stated previously, modelling of these a priori probabilities becomes difficult if no training samples are used, especially if the assumption of a normal distribution of the data vectors is unrealistic, e.g. for built-up areas [Gorte, 1999]. An example of combining various cues in a Bayesian network with the goal of detecting buildings was presented in [Brunn and Weidner, 1997]. A hierarchical strategy is used, turning the classification results of the coarser resolution into one of the cues for the classification in the next iteration. The model of the conditional probabilities relating the classification results is heuristic, which reduces the statistical soundness of the probabilistic approach. We also doubt that the DSM heights of the building roofs are normally distributed, but rather expect a mixture of several normal distributions, each corresponding to a specific building type.

Dempster-Shafer theory of evidence was introduced as an expansion of the probabilistic approach that can also handle imprecise and incomplete information as well as conflict within the data [Klein, 1999, Lee et al., 1987, Le Hégarat-Mascle et al., 1997]. A description of the advantages of the Dempster-Shafer theory, compared to probabilistic reasoning, is given in [Le Hégarat-Mascle et al., 1997]. An important property of that theory is its capability to handle the union of classes. In [Le Hégarat-Mascle et al., 1997], it was applied to unsupervised classification of optical and SAR images. In the context of building detection, the Dempster-Shafer theory has been applied for the final classification of building candidate regions, combining cues such as the NDVI and the average relative heights to distinguish buildings from other objects [Lu and Trinder, 2003].

Even though several authors assess the advantages of the theory of Dempster-Shafer for data fusion in classification [Klein, 1999, Lee et al., 1987, Le Hégarat-Mascle et al., 1997], to our knowledge, that theory has not yet been applied to a per-pixel classification of high resolution remotely sensed data of different origin with the goal of detecting individual buildings. In this chapter, we will show how this can be accomplished for the fusion of LIDAR data and multi-spectral images without using training areas or assigning thematic labels to classes of feature space interactively. The Dempster-Shafer theory is applied in two stages of the overall process, first to detect the candidate building regions and then to eliminate false building candidates. We will describe a heuristic model for the distribution of the evidence provided by our cues to the classes of the classification problem, exploiting the fact that the Dempster-Shafer theory can handle the union of classes, which, in accordance with [Le Hégarat-Mascle et al., 1997], we consider to be a good alternative to other classification techniques for handling the "mixed pixel" problem. We will also evaluate our method using reference data, showing that our method gives satisfactory results in an area of very inhomogeneous building shapes, and that the detectability of buildings mainly depends on the building size, given the resolution of the LIDAR data.

4.2 Fusing LIDAR Data and Multi-spectral Images for Building Detection

4.2.1 Overview of the Process Flow

The input to our method is given by three data sets that have to be generated from the raw data by preprocessing. The *DSM corresponding to the last pulse data* (DSM_L) is a regular height grid interpolated by linear prediction using a straight line as the covariance function, thus almost without filtering. This is accomplished by using the program `SCOP` developed at Vienna University of Technology [Rottensteiner and Briese, 2002]. The first pulse data are also sampled into a regular grid, and by computing the height differences of the first and the last pulse DSMs, a grid ΔH_{FL} of the *height differences between the first and the last pulses* is obtained. The $NDVI$ is computed from the near infrared and the red bands of the multi-spectral images [Lu and Trinder, 2003]. The image data must be geocoded so that the data are already aligned for the subsequent processes.

The work flow for building detection consists of two stages. First, a coarse DTM is generated from the input data. This DTM is used to compute a normalised DSM, which along with parameters of surface roughness of the DSM_L, the $NDVI$, and the height differences ΔH_{FL}, provides the input for the second stage, the detection of building regions by Dempster-Shafer fusion. As the model we use for surface roughness is an important component of our method, it will be described in section 4.2.2.

Coarse DTM generation is based on the hierarchical application of morphological grey scale opening using structural elements of different sizes to overcome the problems caused by large buildings in the data set. After morphological opening, a rule-based algorithm is used to detect large buildings in the data. That information is used in the next iteration, when a smaller structural element is used for morphological opening, to eliminate large buildings. DTM heights computed from the previous iteration are substituted for the results of the morphological filter. The process is finished when the minimum size for the structural element is reached. Our method for hierarchical DTM generation is described in more detail in section 4.2.3. Note that the classification performed here was originally used for building detection alone [Rottensteiner et al., 2003]. As this rule-based algorithm consists of a sequence of thresholding operations, all the available information is never evaluated jointly. Classification errors committed in one of the thresholding operations cannot be corrected in the subsequent stages of the algorithm. Thus, we use this algorithm only to eliminate large buildings in DTM generation (when it is also simple to select some of the thresholds because large buildings are usually characterised by large roof planes). The more sophisticated Dempster-Shafer fusion technique is then used in the final stages of building detection.

The work flow for building detection based on Dempster-Shafer fusion is presented in figure 4.1. First, the normalised DSM (nDSM) and two parameters describing the roughness of the DSM_L, namely the strength and the directedness of surface roughness (cf. section 4.2.2) are computed. Altogether, there are five data sets that contribute to a Dempster-Shafer fusion process carried out for each pixel of the DSM_L grid independently. Each pixel is assigned to one of four classes, namely *building (B)*, *tree (T)*, *grassland (G)*, and *bare soil (S)*. In the subsequent steps, the binary image of the building pixels is used. Morphological filtering helps to eliminate small areas of pixels erroneously classified as building pixels. Then, connected components of building pixels are sought, which results in initial building regions. As the first fusion step accounts for only a very small local neighbourhood (for the evaluation of surface roughness), we eliminate spurious initial building

regions in a second Dempster-Shafer fusion process which utilizes the average NDVI, the average nDSM heights, and two additional attributes derived from the surface roughness parameters. Building detection by data fusion based on the theory of Dempster-Shafer is described in section 4.2.4.

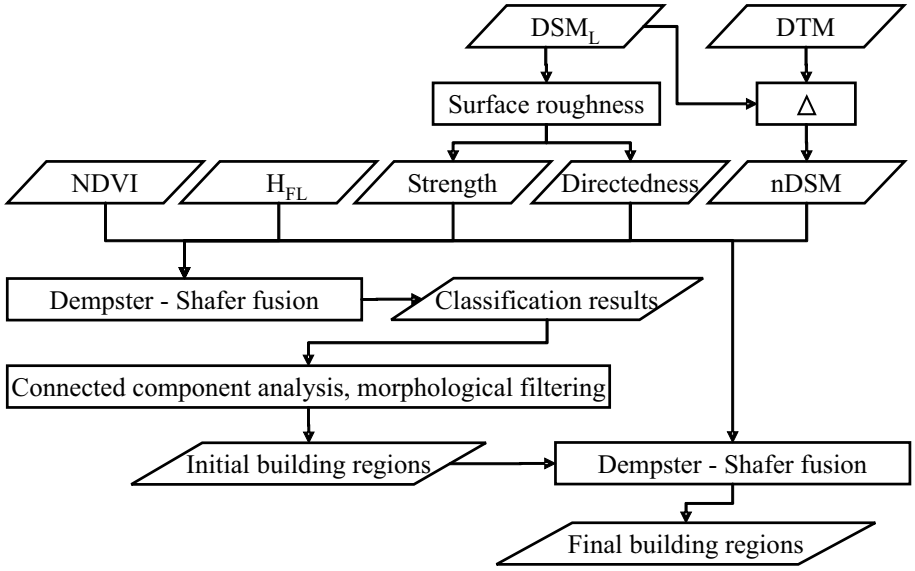


Figure 4.1: The work flow for building detection by data fusion based on the theory of Dempster-Shafer.

4.2.2 Surface Roughness

In this section, the model for DSM surface roughness, which is based on the framework of polymorphic feature extraction [Förstner, 1994, Brügelmann and Förstner, 1992], is presented. In polymorphic feature extraction, the digital image is assumed to consist of regions of homogeneous grey level vectors ("segments"), line regions, and point regions, the latter two being a result of blurring effects in the sensor. The grey level vectors $\mathbf{g}(x,y) = [g_1(x,y),\ldots,g_K(x,y)]^T$ are sampled in a grid (x,y); K represents the number of bands of the image. The grey levels $g_i(x,y)$ of band i are modelled assuming they are affected by additive noise $n_i(x,y)$. For digital images, the noise is assumed to be Poisson-distributed, but the distribution is approximated by a normal distribution $N(0,\sigma_{ni}^2)$ with a signal-dependent noise variance σ_{ni}^2 [Brügelmann and Förstner, 1992]. For our application, σ_{ni}^2 corresponds to the variance of height differences, which can be assumed to be Gaussian.

The first task in polymorphic feature extraction is the classification of each pixel according to whether it is situated in a homogeneous region, a line region, or a point region. With Δg_{ix} and Δg_{iy} denoting the first derivatives of the grey levels of band i by x and y respectively, a matrix $\mathbf{N}$ is computed:

$$\mathbf{N} = \sum_{i=1}^{K} \frac{1}{\overline{\sigma}_{ni}^2} \cdot \mathbf{L} \star \begin{pmatrix} \Delta g_{ix}^2 & \Delta g_{ix} \cdot \Delta g_{iy} \\ \Delta g_{ix} \cdot \Delta g_{iy} & \Delta g_{iy}^2 \end{pmatrix} \tag{4.1}$$

In equation 4.1, $\mathbf{L}$ is a lowpass filter by which the elements of the matrix are convolved. $\overline{\sigma}_{ni}^2$ is the variance of the smoothed grey level differences $\Delta\overline{g}_{ix} = \mathbf{L} \star \Delta g_{ix}$ and $\Delta\overline{g}_{iy} = \mathbf{L} \star \Delta g_{iy}$, which can

be derived from σ_{ni}^2 by error propagation [Förstner, 1994]. Using the elements of $\mathbf{N}$ and denoting the eigenvalues of $\mathbf{N}$ by λ_1 and λ_2, a measure R of homogeneity or *texture strength* and a measure D for *texture directedness* can be defined [Förstner, 1994, Brügelmann and Förstner, 1992]:

$$R = tr(\mathbf{N}) = \sum_{i=1}^{K} \frac{\mathbf{L} \star (\Delta g_{ix}^2 + \Delta g_{iy}^2)}{\overline{\sigma}_{ni}^2} \tag{4.2}$$

$$D = 1 - \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2 = \frac{4 \cdot det(\mathbf{N})}{tr^2(\mathbf{N})} \tag{4.3}$$

Using these two measures, the classification of each pixel is performed. Using a threshold R_{min} for texture strength, a pixel is classified as being in a homogeneous region if $R \leq R_{min}$. If $R > R_{min}$, the local neighbourhood (defined by the extents of the filter $\mathbf{L}$) contains significant grey level variations, i.e., variations of a size that can no longer be explained by noise with a variance σ_{ni}^2. Texture directedness is used to decide whether these variations are isotropic or not. Using a threshold D_{min}, the distribution of the directions of the gradient vectors is considered to be isotropic if $D > D_{min}$, and the pixel is classified as being in a point-like region. If $D \leq D_{min}$, the gradient vectors are supposed to be more or less parallel, which indicates that the pixel belongs to a line region.

The selection of the threshold R_{min} is very critical. If good estimates for σ_{ni}^2 are available, e.g. derived in the way described in [Brügelmann and Förstner, 1992], R can be assumed to follow a χ^2 distribution, and R_{min} is selected relative to the significance level of a hypotheses test [Förstner, 1994]. In some occasions it might be convenient to select

$$R_{min} = j \cdot median(R) \tag{4.4}$$

By relating R_{min} to the median of R, the selection of that threshold is replaced by the selection of a multiplication constant j. Again, this is not too critical because R_{min} is adaptive to the image content (equation 4.4). The selection of the threshold D_{min} for texture directedness is less critical because D is always between 0 and 1. D_{min} can be chosen to be between 0.5 and 0.7 [Förstner, 1994].

We apply the classification scheme of polymorphic feature extraction to the first derivatives of the DSM. Assuming the DSM to be represented by a height grid $z(x, y)$, we obtain a digital "image" of two bands. The grey levels of this image are the first derivatives of $z(x, y)$ by x and y, respectively: $\mathbf{g}(x, y) = [\partial z(x, y)/\partial x, \partial z(x, y)/\partial y]^T$. Under these assumptions, the matrix $\mathbf{N}$ can be computed from:

$$\mathbf{N} = \frac{1}{\overline{\sigma}_x^2} \cdot \mathbf{L} \star \begin{pmatrix} \left(\frac{\partial^2 z}{\partial x^2}\right)^2 & \left(\frac{\partial^2 z}{\partial x^2}\right) \cdot \left(\frac{\partial^2 z}{\partial x \partial y}\right) \\ \left(\frac{\partial^2 z}{\partial x^2}\right) \cdot \left(\frac{\partial^2 z}{\partial x \partial y}\right) & \left(\frac{\partial^2 z}{\partial x \partial y}\right)^2 \end{pmatrix} + \frac{1}{\overline{\sigma}_y^2} \cdot \mathbf{L} \star \begin{pmatrix} \left(\frac{\partial^2 z}{\partial y \partial x}\right)^2 & \left(\frac{\partial^2 z}{\partial y^2}\right) \cdot \left(\frac{\partial^2 z}{\partial y \partial x}\right) \\ \left(\frac{\partial^2 z}{\partial y^2}\right) \cdot \left(\frac{\partial^2 z}{\partial y \partial x}\right) & \left(\frac{\partial^2 z}{\partial y^2}\right)^2 \end{pmatrix} \tag{4.5}$$

In equation 4.5, $\overline{\sigma}_x^2$ and $\overline{\sigma}_y^2$ are the variances of the smoothed matrix elements. They can be derived from an estimate of the variance σ_z^2 of the DSM heights by error propagation. The "grey levels" of the original image are the components of the surface normal vectors. The elements of $\mathbf{N}$ are derived from the second derivatives of the DSM and thus related to the local curvature of the DSM. "Homogeneous" pixels are situated in neighbourhoods of homogeneous surface normal vectors and thus in neighbourhoods of small second derivatives (equations 4.2 and 4.5). Texture strength R can

thus be interpreted as a measure of local co-planarity, or as the strength of local surface roughness. "Line" pixels are situated in neighbourhoods with large, but anisotropic variations of the surface normal vectors. They correspond to surface discontinuities and surface intersection curves. Point pixels occur in areas with large amplitude, isotropic variations of the surface normal vectors, which is typical for building corners and for trees. In the subsequent sections, we will see how texture strength and texture directedness, as well as the results of texture classification are used as cues for building detection.

In our implementation, we choose $\mathbf{L}$ to be a Binomial filter kernel of size $n \times n$ and select R_{min} according to equation 4.4. Thus, there are three control parameters: the size n of the filter kernel, describing the extent of the neighbourhood for which texture classification is carried out, the multiplication constant j from equation 4.4, and D_{min}. We choose n in accordance with the minimum linear extent of a roof plane we expect to be detectable, given the resolution of the DSM (e.g., n = 3). D_{min} is chosen to be 0.7. The most critical choice is the value of j. Typically, we select j to be between 1.0 and 2.0.

4.2.3 Hierarchical DTM Generation

In this section, the process of DTM generation is discussed in detail. For building detection, the DTM should be a relatively good approximation of the terrain, so that the nDSM reflects the actual building heights. If the DTM is to be created by morphological filtering, the size of the structural element must reflect the size of the largest building available in the data set. That is, large structural elements are required in the presence of large buildings. However, this means that terrain structures smaller than the structural element, as well as large buildings, will be eliminated by morphological filtering. Wherever this occurs, the heights of the nDSM will be systematically distorted, because by "cutting off" the tops of small hills the heights of the nDSM may become too large. This will cause errors in classification algorithms when the nDSM is taken as an input. As morphological filtering using a structural element smaller than the largest buildings in the data set will not filter out these buildings, they must be detected and then eliminated from the DTM in the hierarchical procedure outlined earlier in section 4.2.1. In each of the iterations, initial candidate regions for large buildings are detected by thresholding operations, and then surface roughness is evaluated to eliminate regions corresponding to trees.

4.2.3.1 Detecting Candidate Regions for Large Buildings

The morphological filtering provides a coarse approximation for the DTM. In all iterations except the first one, the DTM heights generated in the previous iteration are substituted for the results of morphological filtering in areas where buildings have been detected. This substitution ensures that large buildings that are preserved by morphological filtering are eliminated in the current iteration. Hence, an initial building mask is created by thresholding the heights of the nDSM. This initial building mask still contains singular pixels with large nDSM heights, areas covered by vegetation, and terrain structures smaller than the structural element for morphological filtering. Pixels having an $NDVI$ greater than a certain threshold or a height difference ΔH_{FL} greater than an appropriate threshold are considered to be covered by vegetation and thus erased from the building mask. Also, a binary morphological opening filter using a small structural element is applied to the initial building mask to remove oddly shaped objects and to separate building regions just bridged by a thin line of pixels. The initial building regions are obtained by a connected component analysis of the resulting image.

Small regions are discarded, because at this stage, we only want to detect large buildings in order to eliminate them from the DTM.

4.2.3.2 Classification of Building Candidate Regions

Some of the initial building regions correspond to groups of trees or to small terrain structures. These regions can be eliminated by evaluating a surface roughness criterion derived by an analysis of the second derivatives of the DSM_L, using the method described in section 4.2.2. The numbers of "homogeneous" and "point-like" pixels are counted in each initial building region. Buildings are characterised by a large percentage of "homogeneous" and by a small percentage of "point-like" pixels. By comparing these percentages to thresholds, non-building regions can be eliminated. The surface roughness criterion performs well for large buildings and with dense LIDAR data [Rottensteiner and Briese, 2002]. However, if the point distance of the LIDAR data is larger, e.g. 1 m, only a few LIDAR points are situated on small buildings, so that the percentage of "homogeneous" pixels is reduced, while the percentage of "point-like" pixels is increased, causing the detection of small buildings to be more difficult.

4.2.4 Building Detection Based on Dempster-Shafer Fusion

The Dempster-Shafer theory of evidence is frequently applied for the fusion of data from multiple sensors. Unlike Bayesian probabilistic reasoning, it offers tools to represent partial knowledge about a sensor's contribution to the classification process. We provide an overview on that theory based on [Klein, 1999, Lee et al., 1987, Le Hégarat-Mascle et al., 1997] in section 4.2.4.1. In section 4.2.4.2, we present some general considerations with respect to the definition of the probability masses. Then, we describe the application of the Dempster-Shafer theory in building detection in sections 4.2.4.3 and 4.2.4.4. Note that in section 4.2.4.1, we use the term "sensor" in the way it is done in the cited literature, whereas in the other sections, we rather use the term "cue", because some of the cues we use are derived from one sensor only.

4.2.4.1 Theory of Dempster-Shafer Fusion

Let us assume a classification problem where the input data are to be classified into n mutually exclusive classes. The set θ of these classes is called the *frame of discernment*. The power set of θ is denoted by 2^θ. It contains both the classes and all their possible unions. In the theory of Dempster and Shafer, a *probability mass* $m(A)$ is assigned to every class $A \in 2^\theta$ by a sensor (a cue for classification) such that $0 \leq m(A) \leq 1$, $m(\emptyset) = 0$, and $\sum_{A \in 2^\theta} m(A) = 1$, with $\emptyset$ denoting the empty set.

Imprecision of knowledge can be handled by assigning a non-zero probability mass to the union of two or more classes. Two parameters, support $Sup(A)$ and plausibility $Pls(A)$, can be defined for all $A \in 2^\theta$:

$$Sup(A) = \sum_{B_S \subseteq A} m(B_S) \tag{4.6}$$

$$Pls(A) = \sum_{B_{PL} \cap A \neq \emptyset} m(B_{PL}) = 1 - Sup(\overline{A}) \tag{4.7}$$

In equation 4.6, the sum is taken over all classes $B_S \in 2^\theta$ with $B_S \subseteq A$. The sum in equation 4.7 is taken over all $B_{PL} \in 2^\theta$ with $B_{PL} \cap A \neq \emptyset$. The support of a class is the sum of all masses directly assigned to that class by a data source, whereas the plausibility sums all masses not assigned to the complement of a class. An uncertainty interval $[Sup(A), Pls(A)]$ with $Sup(A) \leq Pls(A)$ can be defined; its length is a measure of the imprecision of knowledge about the uncertainty of class A [Le Hégarat-Mascle et al., 1997]. $\overline{A}$ is the complementary hypothesis of A: $A \cup \overline{A} = \theta$ and $A \cap \overline{A} = \emptyset$. Its support $Sup(\overline{A})$ represents the degree to which the evidence contradicts a proposition. It is called *dubiety*.

If p data sources are available, probability masses $m_i(B_j)$ have to be defined for each data source i with $1 \leq i \leq p$ and for all classes $B_j \in 2^\theta$. The Dempster-Shafer theory allows the combination of these probability masses from several data sources to compute a combined probability mass for each class $A \in 2^\theta$:

$$m(A) = \frac{\sum_{B_1 \cap B_2 \ldots \cap B_p = A} \left(\prod_{1 \leq i \leq p} m_i(B_j) \right)}{1 - \sum_{B_1 \cap B_2 \ldots \cap B_p = \emptyset} \left(\prod_{1 \leq i \leq p} m_i(B_j) \right)} \tag{4.8}$$

The sum in the denominator of equation 4.8 is a measure of the *conflict* in the evidence. As soon as the combined probability masses $m(A)$ have been derived from the original ones, $Sup(A)$, $Pls(A)$, and $Sup(\overline{A})$ can be computed. Finally, a decision rule must be defined in order to determine the accepted simple hypothesis $C_a \in \theta$. There are several ways of defining such a decision rule: C_a can be chosen to be the simple hypothesis (1) of maximum support, (2) of maximum plausibility, or (3) of maximum support without overlapping of uncertainty intervals [Le Hégarat-Mascle et al., 1997]. We use the rule of maximum support in our application.

4.2.4.2 Definition of the Probability Masses

The definition of the probability masses is the distinguishing feature of any application of the Dempster-Shafer theory. There are different strategies on how they can be defined. In [Lee et al., 1987], non-zero probability masses derived from probabilities from a previous classification are assigned to the simple hypotheses, and the only compound class receiving a non-zero probability mass is θ, which is used to model the imprecision of the initial classifications. In [Le Hégarat-Mascle et al., 1997] it is argued that such a definition is appropriate in cases where the information provided by the different sources is mainly redundant. The authors propose two strategies for assigning the probability masses in cases where the information from the different sources is complementary, which is particularly useful if two classes C_i and C_j cannot be distinguished by a certain sensor. In this case, a non-zero probability mass should be assigned to $C_i \cup C_j$. Thus, $m(C_i \cup C_j) \neq 0$. C_i and C_j can either be assigned a zero probability mass, thus $m(C_i) = m(C_j) = 0$, or the same mass as $C_i \cup C_j$, thus $m(C_i \cup C_j) = m(C_i) = m(C_j) \neq 0$. The first assignment assumes total ignorance about the membership of a pixel to either C_i or C_j. In [Le Hégarat-Mascle et al., 1997], the second strategy is preferred for practical reasons: the paper deals with unsupervised classification, and applying the first strategy resulted in too many clusters in feature space.

We prefer the first strategy. Unlike the authors of [Le Hégarat-Mascle et al., 1997], we do know the thematic classes we want to distinguish from the beginning, and their number is small. As we shall see in the subsequent sections, the cues (sensors) we use can distinguish two complementary subsets $B_1 \in 2^\theta$ and $B_2 \in 2^\theta$, with $B_1 \cap B_2 = \emptyset$ and $B_1 \cup B_2 = \theta$. Both B_1 and B_2 consist of simple

classes $C_{gh} \in \theta$ with $B_1 = \{C_{11} \dots C_{m1}\}$ and $B_2 = \{C_{12} \dots C_{k2}\}$, m and k being the number of simple classes in B_1 and B_2, respectively. Of course, B_1 and B_2 can be different for different cues; otherwise, the information provided by the sensors would not be complementary. If neither B_1 nor B_2 consists of one simple class only, our ignorance about the simple classes C_{gh} is complete, and we can see no reason why we should assign a non-zero probability mass to any of them. Non-zero probability masses are thus only assigned to B_1 and B_2. Further, as B_1 and B_2 are mutually exclusive, $m(B_2) = 1 - m(B_1)$ holds true. As a consequence, the problem of defining probability masses is reduced to defining a probability mass function $P_i(x)$ for B_1 for each sensor i with $m_i(B_1) = P_i(x)$, where x is the output of sensor i. $P_i(x)$ can be interpreted as the probability of a certain pixel or feature to belong to class B_1 given that the output of sensor i equals x. It can also be interpreted as the result of an initial classification using only sensor i, and distinguishing only classes B_1 and B_2. In the following sections, we shall describe how the probability mass functions $P_i(x)$ are defined.

4.2.4.3 Initial Land Cover Classification

In this process, we want to achieve a per-pixel classification of the input data into one of four classes: Buildings (B), trees (T), grassland (G), and bare soil (S). Five cues derived from the original input data are used for this purpose. None of the five cues could be used individually to distinguish sharply between the four classes B, T, G, and S. They are considered to be five data sources in the Dempster-Shafer fusion process. Table 4.1 shows our definition of the initial probability masses according to the guidelines developed in section 4.2.4.2 and the way they propagate to the final probability masses:

- The height differences ΔH between the DSM_L and the DTM can be used to distinguish elevated objects from the ground. In our classification scheme, the trees and the buildings are elevated objects, whereas the other classes represent the ground. We assign a probability mass $P_{\Delta H} = P_{\Delta H}(\Delta H)$ to the combined class $B \cup T$, and $1 - P_{\Delta H}$ to $G \cup S$. $P_{\Delta H}$ is chosen to be an increasing function of ΔH.

- The strength R of surface roughness grows with the second derivatives of the DSM (cf. equations 4.2 and 4.5), thus with changes of the surface normal vectors. Large variations of the surface normal vectors are typical for trees, the only object class in our classification scheme not having a smooth surface, so that we use R to distinguish trees from other object classes. We assign a probability mass $P_R = P_R(R)$ to class T , and $1 - P_R$ to $B \cup G \cup S$. By that assignment, we neglect the fact that R can also be very large at the building boundaries and at the step edges of the terrain. P_R is chosen to be an increasing function of R.

- The directedness D of surface roughness depends on the local variations of the directions of the surface normal vectors. With trees, surface normal vectors usually do not change in an isotropic way. That is why D is also an indicator for trees, but only if R is above a certain threshold (because otherwise, D might be dominated by the noise in a planar area). We assign a probability mass $P_D = P_D(D, R)$ to class T, and $1 - P_D$ to $B \cup G \cup S$. P_D is chosen to be an increasing function of D if R is greater than a certain threshold.

- The height differences ΔH_{FL} between the first and the last pulse DSMs are used in a similar way as the strength R of surface roughness. We neglect the large values of ΔH_{FL} at building boundaries and at power lines. We assign a probability mass $P_{FL} = P_{FL}(\Delta H_{FL})$ to class T, and $1 - P_{FL}$ to $B \cup G \cup S$. P_{FL} is chosen to be an increasing function of ΔH_{FL}.

- The $NDVI$ is an indicator for vegetation, thus for classes T and G. A probability mass $P_N = P_N(NDVI)$ is assigned to the combined class $T \cup G$, and $1 - P_N$ to $B \cup S$. P_N is chosen to be an increasing function of the $NDVI$.

For modelling the probability masses $P_i(x)$ for these five cues, we use a heuristic approach that shares some similarities to the concept of membership functions in fuzzy logic [Vögtle and Steinle, 2003]. We have described previously why each of the cues separates two subsets of θ in our classification scheme. For input parameters x smaller than a threshold x_1, we assume the assignment of a pixel to class B_1 (cf. section 4.2.4.2) to be very unlikely, which is modelled by a small probability mass P_1. For input parameters above a second threshold x_2 with $x_1 < x_2$, we assume the assignment of a pixel to class B_1 to be almost certain, which is modelled by a rather large probability mass P_2, with $0 \leq P_1 < P_2 \leq 1$. For instance, if ΔH is smaller than 1.5 m, it is very unlikely that there is a building or a tree; if it is greater than 3 m, it is very unlikely that there is grassland or bare soil. Between x_1 and x_2, the probability mass function should be defined in a way that there are no step edges (which would correspond to applying "hard" thresholds), but rather a smooth transition between the two probability levels P_1 and P_2. This could be a straight line, but we use a cubic parabola with horizontal tangents (thus being differentiable) at $(x = x_1)$ and at $(x = x_2)$ (cf. figure 4.2). Thus, the probability masses $P_{\Delta H}$, P_R, P_{FL}, and P_N are computed according to equation 4.9:

Class	ΔH	R	D	ΔH_{FL}	$NDVI$	Combined probability mass
B	0	0	0	0	0	$\frac{P_{\Delta H} \cdot (1-P_R) \cdot (1-P_D) \cdot (1-P_{FL}) \cdot (1-P_N)}{1-C}$
T	0	P_R	P_D	P_{FL}	0	$\frac{P_{\Delta H} \cdot P_R \cdot P_D \cdot P_{FL} \cdot P_N}{1-C}$
G	0	0	0	0	0	$\frac{P_N \cdot (1-P_{\Delta H}) \cdot (1-P_R) \cdot (1-P_D) \cdot (1-P_{FL})}{1-C}$
S	0	0	0	0	0	$\frac{(1-P_{\Delta H}) \cdot (1-P_R) \cdot (1-P_D) \cdot (1-P_{FL}) \cdot (1-P_N)}{1-C}$
$B \cup T$	$P_{\Delta H}$	0	0	0	0	0
$B \cup G$	0	0	0	0	0	0
$B \cup S$	0	0	0	0	$1 - P_N$	0
$T \cup G$	0	0	0	0	P_N	0
$T \cup S$	0	0	0	0	0	0
$G \cup S$	$1 - P_{\Delta H}$	0	0	0	0	0
$B \cup T \cup G$	0	0	0	0	0	0
$B \cup T \cup S$	0	0	0	0	0	0
$B \cup G \cup S$	0	$1 - P_R$	$1 - P_D$	$1 - P_{FL}$	0	0
$T \cup G \cup S$	0	0	0	0	0	0
θ	0	0	0	0	0	0

Table 4.1: The probability masses for the initial classification. Classes: buildings (B), trees (T), grassland (G), and bare soil (S). ΔH: initial probability masses for the height differences in the normalised DSM_L. R: initial probability masses for the strength of surface roughness. D: initial probability masses for the directedness of surface roughness. ΔH_{FL}: initial probability masses for the height differences between first and last pulses. $NDVI$: initial probability masses for the $NDVI$. The conflict C is the sum in the denominator of equation 4.8:

$$\begin{aligned} C = {} & P_R \cdot (1 - P_{\Delta H} \cdot P_D) + P_D \cdot (1 - P_R) + P_{\Delta H} \cdot P_R \cdot P_D \cdot (1 - P_N) + \\ & + P_{\Delta H} \cdot P_N \cdot (1 - P_R) \cdot (1 - P_D) + P_{\Delta H} \cdot P_{FL} \cdot (1 - P_R) \cdot (1 - P_D) \cdot (1 - P_N) + \\ & + P_{FL} \cdot P_N \cdot (1 - P_{\Delta H}) \cdot (1 - P_R) \cdot (1 - P_D) + P_{FL} \cdot (1 - P_{\Delta H}) \cdot (1 - P_R) \cdot (1 - P_D) \cdot (1 - P_N) + \\ & + P_{\Delta H} \cdot P_R \cdot P_D \cdot P_N \cdot (1 - P_{FL}) \end{aligned}$$

$$P_i(x) = \begin{cases} P_1 & \forall x | x \leq x_1 \\ P_1 + (P_2 - P_1) \cdot \left[3 \cdot \left(\frac{x - x_1}{x_2 - x_1} \right)^2 - 2 \cdot \left(\frac{x - x_1}{x_2 - x_1} \right)^3 \right] & \forall x | (x > x_1) \wedge (x < x_2) \\ P_2 & \forall x | x \geq x_2 \end{cases} \tag{4.9}$$

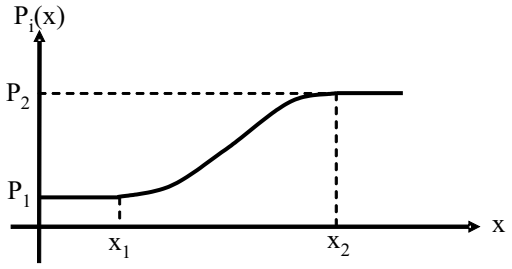


Figure 4.2: The probability mass function.

A slightly different definition has to be used for the probability mass function P_D because P_D is only significant if R is significant also. Thus, if R is below a threshold R_{min}, P_D cannot contribute to the classification, which is modelled by assigning a probability mass of 0.5 to both T and $B \cup G \cup S$. Otherwise, the probability mass is also modelled by equation 4.9. We choose $R_{min} = 5 \cdot median(R)$. By the latter selection, the threshold is made adaptive to the average surface roughness of a scene (cf. section 4.2.2).

We use $P_1 = 5\%$ and $P_2 = 95\%$. Note that $P_1 \neq 0\%$ and $P_2 \neq 100\%$. This is the main difference between our functional model and the membership functions for fuzzy logic described in [Vögtle and Steinle, 2003]. We never assume the information from a sensor to be 100% certain. Conflicts in sensor information are expected both because the number of classes in our classification scheme is much smaller than the number of object classes that are actually observable in the data and the sensor data exhibit random variation. Actually, unlike other classification techniques, the Dempster-Shafer theory gives a direct measure of that conflict. If two sensors contradict each other and if the information of both sensors are considered to be 100% certain, that conflict cannot be resolved: the conflict would be equal to 1.0, and equation 4.8 could not be evaluated. We do not consider this to be a restriction of the Dempster-Shafer theory, but rather consider it to be an advantage that such a situation can be clearly detected.

The values we use for (x_1, x_2) are listed in table 4.2. As with the value for P_D, they were determined empirically, but are assumed to be generally applicable. The values (x_1, x_2) for P_R are linked to the median of R and thus adaptive to the average slope variations in the data. The combined probabilities are evaluated for each pixel independently, and the pixel is assigned to the class of maximum support. In comparison to the rule-based algorithm described in [Rottensteiner et al., 2003] which is used for hierarchical DTM generation, there are several improvements.

- All cues are evaluated simultaneously.
- No sharp thresholds are required, and the probability mass function (equation 4.9) has a smooth transition between the two levels P_1 and P_2.
- We do not eliminate parts of the data during the process as not belonging to the class "building", but we achieve an overall classification of land cover with respect to our four classes.

	$\Delta H[m]$	R	D	$\Delta H_{FL}[m]$	$NDVI[\%]$
x_1	1.5	$2 \cdot median(R)$	0.1	1.5	30
x_2	3.0	$15 \cdot median(R)$	0.9	3.0	65

Table 4.2: The values for (x_1, x_2) for $P_{\Delta H}$, P_R, P_D, P_{FL}, and P_N.

4.2.4.4 Final Classification of Building Regions

After the initial classification, we obtain a binary building mask of all pixels classified as "building". As all pixels were classified independently from each other, only a small local neighbourhood contributed to their classification (via R and D), which causes classification errors:

- There are singular "building" pixels inside larger areas of other classes, or "tree" pixels inside building roofs.
- R and D give relatively large values for small detached houses if the resolution of the LIDAR data is not better than or equal to about 0.5 m. This might result in classification errors that cannot be corrected by a local analysis.

To overcome these problems, isolated building pixels are eliminated by binary morphological opening, and a building label image is created by a connected component analysis. After that, the Dempster-Shafer theory is applied for a final classification of the original building regions thus detected, this time using four cues representing average values for each building region (table 4.3):

- The average height differences ΔH_a between the DSM_L and the DTM are used in a similar way as ΔH in the original classification by assigning a probability mass $P_{\Delta Ha} = P_{\Delta Ha}(\Delta H_a)$ to $B \cup T$, and $(1 - P_{\Delta Ha})$ to $G \cup S$.
- The percentage H of pixels classified as "homogeneous" in polymorphic feature extraction (cf. section 4.2.2) is an indicator for objects consisting of planar surface patches. Thus, we assign a probability mass $P_H = P_H(H)$ to class $B \cup G \cup S$, and $(1 - P_H)$ to class T.
- The percentage P of pixels classified as "point-like" in polymorphic feature extraction is an indicator for trees. We assign a probability mass $P_P = P_P(P)$ to class T, and $(1 - P_P)$ to $B \cup G \cup S$.
- The average $NDVI_a$ is used in a similar way as in the original classification by assigning a probability mass $P_{Na} = P_{Na}(NDVI_a)$ to $T \cup G$, and $(1 - P_{Na})$ to $B \cup S$.

The height differences between first and last pulses are no longer used. For the probability masses $P_{\Delta Ha}$, P_H, P_P, and P_{Na} we again use the function described by equation 4.9, with $P_1 = 5\%$ and $P_2 = 95\%$. The values for (x_1, x_2) are presented in table 4.4. These values are designed to avoid eliminating buildings erroneously, but to accept a larger false-alarm rate. The combined probability masses are evaluated for each initial building region, and if such a region is assigned to a class other than "building", it is eliminated. Finally, the regions classified as buildings are grown by a few pixels by morphological filtering to correct for building boundaries being erroneously classified as trees, caused by the large values for R (cf. section 4.2.4.3). The advantage of this method compared to the one described in [Rottensteiner et al., 2003] is that it considers all cues simultaneously and avoids sharp thresholds.

Class	ΔH_a	H	P	$NDVI_a$	Combined probability mass
B	0	0	0	0	$\frac{P_{\Delta Ha} \cdot P_H \cdot (1-P_P) \cdot (1-P_{Na})}{1-C}$
T	0	$1-P_H$	P_P	0	$\frac{P_{\Delta Ha} \cdot (1-P_H) \cdot P_P \cdot P_{Na}}{1-C}$
G	0	0	0	0	$\frac{P_{Na} \cdot (1-P_{\Delta Ha}) \cdot P_H) \cdot (1-P_P)}{1-C}$
S	0	0	0	0	$\frac{(1-P_{\Delta Ha}) \cdot P_H \cdot (1-P_P) \cdot (1-P_{Na})}{1-C}$
$B \cup T$	$P_{\Delta Ha}$	0	0	0	0
$B \cup G$	0	0	0	0	0
$B \cup S$	0	0	0	$1-P_{Na}$	0
$T \cup G$	0	0	0	P_{Na}	0
$T \cup S$	0	0	0	0	0
$G \cup S$	$1-P_{\Delta Ha}$	0	0	0	0
$B \cup T \cup G$	0	0	0	0	0
$B \cup T \cup S$	0	0	0	0	0
$B \cup G \cup S$	0	P_H	$1-P_P$	0	0
$T \cup G \cup S$	0	0	0	0	0
θ	0	0	0	0	0

Table 4.3: The probability masses for the final classification. Classes: buildings (B), trees (T), grassland (G), and bare soil (S). ΔH_a: initial probability masses for the average height of the building. H: initial probability masses for the the percentage of "homogeneous" pixels. P: initial probability masses for the percentage of "point" pixels. $NDVI_a$: initial probability masses for the average $NDVI$. The conflict C is the sum in the denominator of equation 4.8:
$C = (1-P_h) \cdot (1-P_{\Delta Ha} \cdot P_P) + P_P \cdot P_H + P_{\Delta Ha} \cdot P_P \cdot (1-P_H) \cdot (1-P_{Na}) + P_{\Delta Ha} \cdot P_H \cdot P_{Na} \cdot (1-P_P)$

	$\Delta H_a[m]$	$H[\%]$	$P[\%]$	$NDVI_a[\%]$
x_1	1.5	0	30	30
x_2	3.0	60	75	65

Table 4.4: The values for (x_1, x_2) for $P_{\Delta Ha}$, P_H, P_P, and P_{Na}.

4.3 Experiments

4.3.1 Description of the Data Set

The test data set was captured over Fairfield (NSW) using an Optech ALTM 3025 laser scanner. It covers an area of 2 $\times$ 2 km^2. Both first and last pulses and intensities were recorded with an average point distance of about 1.2 m. We derived DSM grids at a resolution of 1 m from these data. A true colour digital orthophoto with a resolution of 0.15 m was also available for the area. The orthophoto had been created using a DTM, so that the roofs and the tree-tops were displaced with respect to the LIDAR data. Thus, data alignment was not perfect. Unfortunately, the digital orthophoto did not contain an infrared band. We circumvented this problem by resampling both the digital orthophoto and the (infrared) LIDAR intensity data to a resolution of 1 m and by computing a "pseudo-NDVI-image" from the LIDAR intensities and the red band of the digital orthophoto. Apart from problems with data alignment caused by the displaced tree canopies in the orthophoto, there were also problems with shadows in the orthophoto, so the pseudo-NDVI-image did not provide as

much information as expected.

A reference data set was created by digitising building polygons interactively in the digital orthophoto. We chose to digitize all structures recognisable as buildings independent of their size. The reference data include garden sheds, garages, etc., that are sometimes smaller than 10 m^2 in area. Such small structures cannot be expected to be detected in the LIDAR data, given their reso-

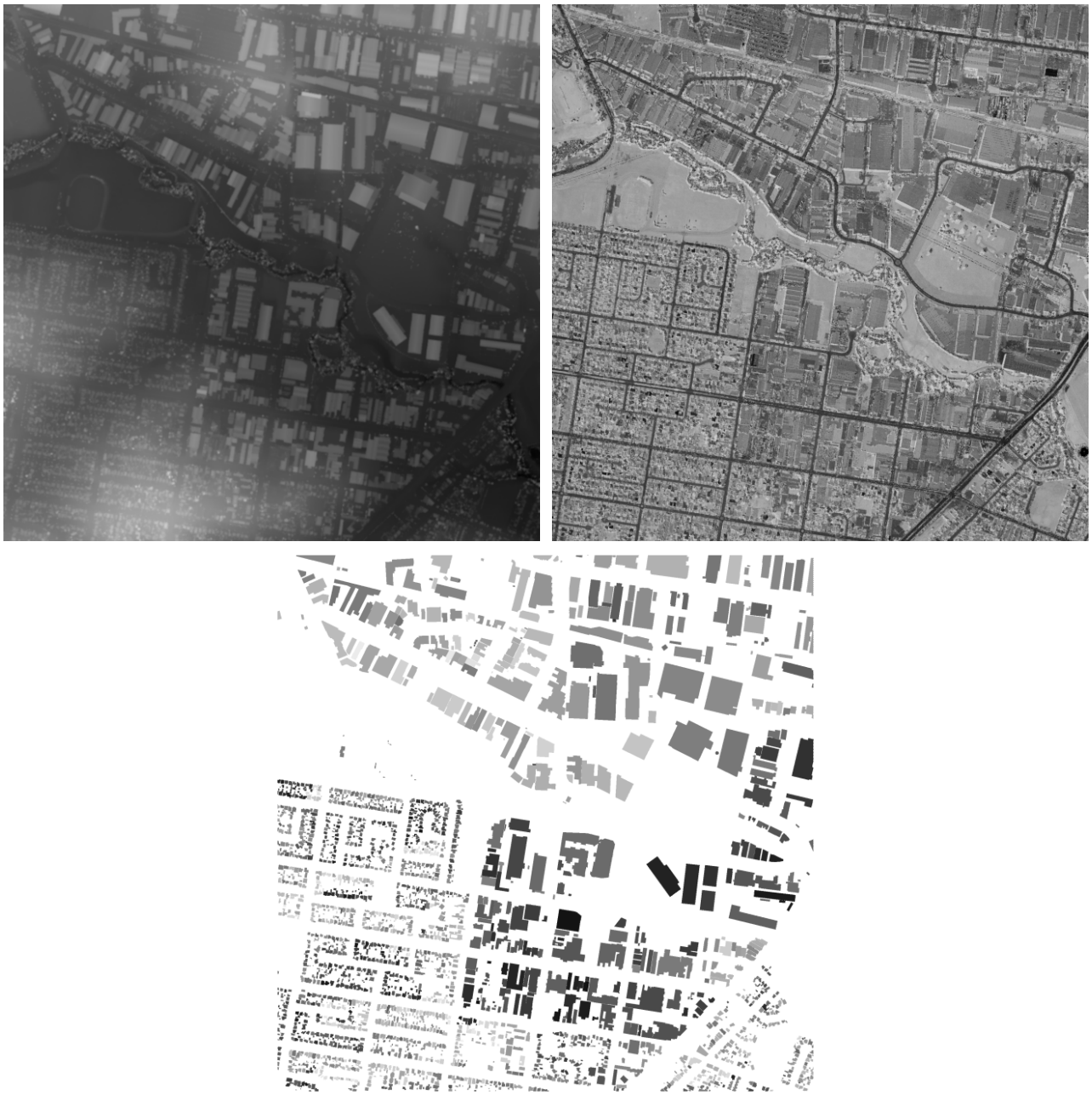


Figure 4.3: The Fairfield data set. First row, left: the DSM from last pulse data (black: low areas, white: high areas). First row, right: the pseudo-NDVI image. Second row: a label image derived from the reference data set. Total area: 2000 $\times$ 2000 m^2

lution. Neighbouring buildings that were joined but are obviously separate entities were digitized as separate polygons. Thus, altogether 2385 polygons were digitized. As the aerial image used for producing the orthophoto and the LIDAR data were captured at different epochs, there were contradictions in the two data sets as some buildings were either constructed or demolished between them. Altogether 49 polygons were detected in only one data set. Of these, all except one (an industrial building) were larger detached houses. Smaller entities could not be checked in the LIDAR data. The DSM and the pseudo-NDVI image as well as a label image created from the reference data set are shown in figure 4.3.

4.3.2 Method of Evaluation

In the evaluation process, we compared two data sets: the "automatic data set" consisting of building regions detected automatically, and the reference data set. The comparison of two spatial data sets is not a straightforward task if their topologies are different [Ragia and Winter, 2000]. This is the case here because the automatic building detection process cannot separate buildings that are actually joined in object space or that are so close to each other that there are no LIDAR points between them. We have to expect buildings to be merged in the detection process. Thus, a comparison of the boundary polygons proposed in [Ragia and Winter, 2000] is not appropriate. Alternatively, we compare two label images: the label image that is the output of automatic building detection, hence called the "automatic label image", and a label image created from the polygons of the reference data set with the same spatial resolution as the automatic label image, hence called the "reference label image". Before the actual evaluation process, the areas covered by the polygons detected to be available in one data set only are erased in both of these label images. For an evaluation of automatic feature extraction using a reference data set, two numbers of interest are the completeness and the correctness of the results [Heipke et al., 1997]:

$$Completeness = \frac{TP}{TP + FN} \tag{4.10}$$

$$Correctness = \frac{TP}{TP + FP} \tag{4.11}$$

In equations 4.10 and 4.11, TP denotes the number true positives, i.e., the number of entities found to be available in both data sets. FN is the number of false negatives, i.e., the number of entities in the reference data set that were not detected automatically, and FP is the number of false positives, i.e., the number of entities that were detected, but do not correspond to an entity in the reference data set. We are interested in determining completeness and correctness for two types of entities. First, we want to derive them on a per-pixel level. In this context, TP is the number of pixels classified as "building" in both label images, FN is the number of building pixels in the reference label image not classified as "building" in the automatic label image, and FP is the number of building pixels in the automatic label image not classified as "building" in the reference label image. The numbers thus derived for completeness and correctness give a balanced estimate of the area that has been correctly classified as "building". Second, we are interested in numbers on a per-building level, showing how many buildings could be detected and how many of the buildings detected automatically did actually correspond to buildings. In this case, TP, FN, and FP cannot be determined easily because of the problem of multiple and partial overlaps of building regions in the automatic and in the reference data sets.

We denote the sets of regions from the automatic and the reference label images by L_a and L_r, respectively. For each co-occurrence of two labels $l_a \in L_a$ and $l_r \in L_r$, we compute the overlap ratios $p_{ar} = n_{a\cap r}/n_a$ and $p_{ra} = n_{a\cap r}/n_r$, where $n_{a\cap r}$ is the number of common pixels assigned to the region l_a in the automatic label image and to l_r in the reference label image, n_a is the total number of pixels assigned to the region l_a in the automatic label image, and n_r is the total number of pixels assigned to l_r in the reference label image. Obviously, if the two data sets were characterised by the same topology, both p_{ar} and p_{ra} would be close to 100% for all building regions, and there would be exactly one region $l_a \in L_a$ corresponding to each $l_r \in L_r$ and vice versa. As this is not the case, we have to evaluate the overlap percentages p_{ar} and p_{ra} further, to match corresponding regions in the two data sets.

Initially, all tuples (l_a, l_r) with $n_{a\cap r} > 0$ are considered to be possible correspondences. Our analysis starts by eliminating spurious correspondences. We define a function $overlap(l_i, l_j)$ classifying the overlap between regions $l_i \in L_i$ and $l_j \in L_j$ with $i \in \{r, a\}$, $j \in \{r, a\}$, and $i \neq j$:

$$overlap(l_i, l_j) \begin{cases} strong & \forall i, j | p_{ij} > 80\% \\ partial & \forall i, j | 80\% \leq p_{ij} < 50\% \\ weak & \forall i, j | 50\% \leq p_{ij} < 10\% \\ none & \forall i, j | p_{ij} \leq 10\% \end{cases} \tag{4.12}$$

The function defined in equation 4.12 is not necessarily symmetric, thus we cannot expect $overlap(l_a, l_r)$ to be identical to $overlap(l_r, l_a)$. We consider a correspondence between two regions $l_a \in L_a$ and $l_r \in L_r$ to be spurious if both $overlap(l_a, l_r)$ and $overlap(l_r, l_a)$ are either $weak$ or $none$. These correspondences are no longer considered. For each region $l_a \in L_a$ we obtain a subset $L_{ar} \subset L_r$ containing all regions from L_r that correspond to l_a:

$$L_{ar} = \{l_r \in L_r \mid [overlap(l_r, l_a) \in \{strong, partial\}] \vee [overlap(l_a, l_r) \in \{strong, partial\}]\} \tag{4.13}$$

In the same way, for each region $l_r \in L_r$ we obtain the subset $L_{ra} \subset L_a$ of corresponding regions from L_a:

$$L_{ra} = \{l_a \in L_a \mid [overlap(l_r, l_a) \in \{strong, partial\}] \vee [overlap(l_a, l_r) \in \{strong, partial\}]\} \tag{4.14}$$

L_{ra} can be interpreted as the set of regions of the automatic data set into which a region l_r of the reference data set is split. L_{ar} is the set of regions of the reference data set which are merged into a region l_a of the automatic data set. Having found corresponding regions, the overall coverage d_i for each region $l_i \in L_i$ can be computed with $i \in \{r, a\}$, $j \in \{r, a\}$, $i \neq j$, and $n_{a\cap r} = n_{r\cap a}$:

$$d_i = \frac{\sum_{l_j \in L_{ij}} n_{i\cap j}}{n_i} \tag{4.15}$$

Thus, d_i is the ratio between the sum of the number of all pixels overlapping with one of the corresponding regions of the other data set and the total number of pixels of region l_i. For a region $l_r \in L_r$, d_r is the percentage of the area of l_r that is substantially covered by regions detected automatically. We consider a region l_r to be *completely detected* if $d_r > 80\%$, *partly detected* if $80\% \leq d_r < 50\%$, *hardly detected* if $50\% \leq d_r < 10\%$, and *not detected* if $d_r \leq 10\%$. For a region $l_a \in L_a$, d_a

is the percentage of the area of l_a that actually corresponds to regions of the reference data set. We consider a region l_a to be *completely correct* if $d_a > 80\%$, *partly correct* if $80\% \leq d_a < 50\%$, *hardly correct* if $50\% \leq d_a < 10\%$, and *not correct* if $d_a \leq 10\%$. This gives us the tools for computing the numbers of TP, FN, and FP in equations 4.10 and 4.11:

- In equation 4.10, TP is the number of building regions in the reference data set that are either *completely detected* or *partly detected*.
- In equation 4.11, TP is the number of building regions in the automatic data set that are either *completely correct* or *partly correct*.
- FN is the number of building regions in the reference data set that are neither *completely detected* nor *partly detected*.
- FP is the number of building regions in the automatic data set that are neither *completely correct* nor *partly correct*.

Note that the different definitions of TP for computing completeness and correctness is a consequence of the fact that we consider regions that are either split or merged to be correct if the overall coverage is sufficient. We assume a region with coverage larger than 80% to be completely detected/correct because we have to consider errors at the building boundaries, which might comprise a considerable percentage of smaller buildings, and we believe that such errors can be corrected in a later stage of processing if the delineations of the roof planes are searched [Rottensteiner and Briese, 2002].

4.3.3 Results

Figure 4.4 shows the normalised DSM created by three iterations of morphological opening, using structural elements of 150, 75, and 25 m. In the second iteration, buildings larger than 225 m^2, containing at least 45% of homogeneous and less than 20% of point-like pixels are detected. They are

Figure 4.4: The normalised DSM used in Dempster-Shafer classification (black: low areas, white: high areas).

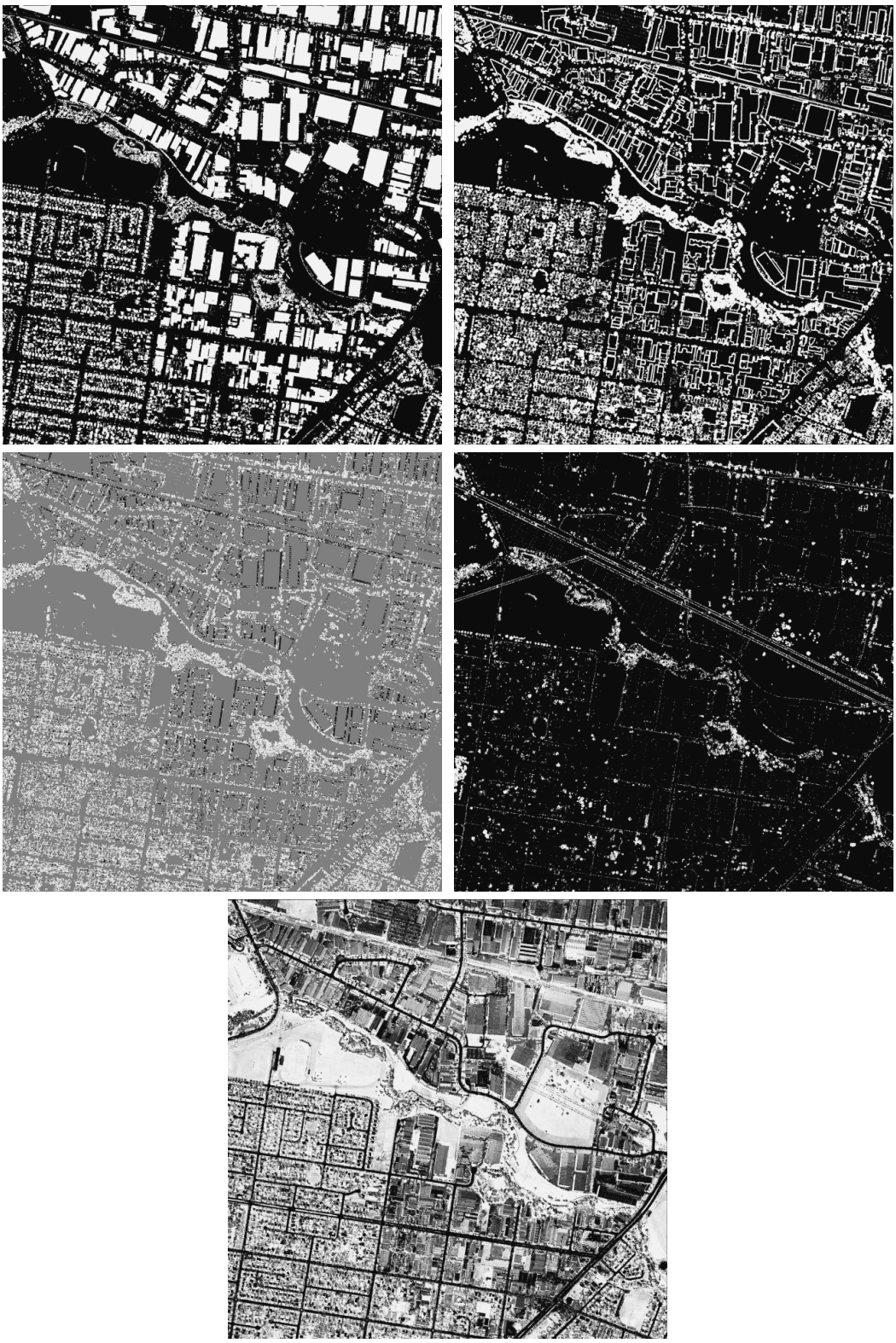


Figure 4.5: The probability masses for the initial Dempster-Shafer classification. First row: $P_{\Delta H}$ (left), P_R (right); second row: P_D (left) and P_{FL} (right); third row: P_N.

eliminated in the third iteration of DTM generation, so that they are preserved in the nDSM in figure 4.4.

Figure 4.5 depicts the probability masses for the initial Dempster-Shafer classification. Note that $P_{\Delta H}$ is large both for buildings and trees, whereas P_R, P_D and P_{FL} give large values for trees. P_R also highlights the building boundaries. The areas in a medium grey in P_D are those where P_D was set to 0.5 because the texture was not considered to be significant. In these areas, P_D did not contribute to the classification. The dominant linear structures in P_{FL} are powerlines, which are not considered by our approach. P_N distinguishes bare soil and, less clearly, buildings from vegetation. Note that the industrial buildings have a relatively large pseudo-NDVI. The results of Dempster-Shafer classification are presented in figure 4.6.

Figure 4.6: The results of the initial Dempster-Shafer classification. White: grass-land (G), light grey: bare soil (S), dark grey: trees (T), black: buildings (B).

After morphological opening of the binary image of the building pixels from figure 4.6 and after eliminating building candidate regions smaller than 10 m^2, the second Dempster-Shafer classification is carried out for altogether 2291 building candidate regions detected in the data. 344 of these regions are found to belong to a class other than "building", so that we finally obtain 1947 building regions. Figure 4.7 shows the final building label image after growing the building regions by morphological closing to compensate for the incorrect classification of the building boundaries. The computation time for achieving the result in figure 4.7 was about 5 minutes on a Pentium 4 PC (2.66 GHz, 512 MB RAM).

Figure 4.7: The final building label image after the second Dempster-Shafer classification and after growing the building regions by morphological closing.

4.3.4 Evaluation of the Results and Discussion

By evaluating the results of the initial classification using the Dempster-Shafer technique in figure 4.6, it is clear that the class "bare soil" mainly corresponds to streets and parking lots. Most of the trees are situated in the valley of the river crossing the scene diagonally, along the streets in the residential area in the south-western part of the scene, and in the backyards of the houses. Step edges at the building boundaries are often classified as trees. Given the resolution of the LIDAR data, it was not easy to separate trees from buildings in the residential areas, which is also the reason for the rugged appearance of the building boundaries in these regions. A few of those residential buildings were erroneously classified as trees, especially if the roof consisted of many small faces. Further problems occurred with bridges because morphological filtering results in a DTM corresponding to the terrain below the bridge. This results in large height differences ΔH that indicate the presence of a building or a tree, whereas the surface roughness of the street and the $NDVI$ indicate an area not covered by trees or vegetation, so that the overall classification would assign such areas to the class "building". There are also isolated spots on top of the large industrial buildings that are not classified correctly, which is mostly caused by chimneys or other objects yielding a large local variation of the surface normal vectors. In the parking lots, there are errors where cars are parked. Of course, cars were not considered in our classification scheme. Some dominant power lines are still partly preserved in the data, classified as "bare soil" inside grassland areas, which is caused by the low reflectance of the power lines in the wavelength of the laser scanner. In general, shadows in the colour orthophoto were an error source, because no shadows appeared in the LIDAR intensity data, so that the "pseudo-NDVI-image" gave systematically wrong $NDVI$ values in these areas. Finally, there are also some incorrectly classified isolated pixels inside larger areas of another class, because context was not considered in the first classification process.

The parameter settings for the second classification seem to be more critical and more dependent on the data than those for the first, because with increasing resolution of the LIDAR data, the percentage of "homogeneous" pixels in roof planes will also increase, whereas the percentage of "point" pixels will become smaller. This implies that the second classification stage will be more helpful for the discrimination of buildings and trees with LIDAR data of a higher resolution than it is in our test project. This claim is confirmed in [Rottensteiner and Briese, 2002]. Figure 4.8 shows the results of the evaluation of the automatic building extraction process. On a per-pixel basis, the completeness was 94%, thus 94% of the building pixels were actually detected, which is very satisfactory. The missed buildings (black areas in figure 4.8) were small residential buildings, some having roofs with high reflectance in the wavelength of the laser scanner (thus, a high pseudo-NDVI), others having roofs consisting of many small planar faces, or they are too small to be detected given the resolution of the LIDAR data. For a few large industrial buildings, some building parts could not be detected due to errors in DTM generation. However all large buildings except one (at the upper margin) were substantially detected. The correctness on a per-pixel basis was 85%, thus 85% of the pixels classified as building pixels do actually correspond to a building. This is not quite as good as the completeness. Figure 4.8 shows that this number is affected by errors at the building boundaries. After enlarging the buildings using the approach described in section 4.2.4.4, the buildings seem to be slightly too large. There are also a few larger false positives at bridges, at small terrain structures not covered by vegetation, in areas with overseas containers, and in trailer parks.

The results of the evaluation on a per-building basis are presented in figure 4.9. The upper diagram shows both completeness and correctness depending on the area covered by the buildings using a class width of 10 m^2, i.e., it shows completeness and correctness computed separately for all

Figure 4.8: Evaluation of the results of building detection. Light grey: correct building pixels; medium grey: false positives; black: false negatives.

area intervals of the abscissa. It is obvious that the quality of the results depends on the building size: for buildings of a size larger than 90 m^2, completeness is greater than 90%, but it becomes less than 50% for buildings smaller than 40 m^2. A similar trend can be observed for the correctness, which is again not as good as completeness. For buildings larger than 130 m^2, correctness is greater than 90%. For buildings between 130 and 40 m^2, correctness oscillates between 90% and 50%, and it rapidly descends toward zero for detected building regions smaller than 40 m^2. The lower diagram in figure 4.9 shows the cumulative completeness and correctness for buildings covering an area larger than the size shown in the abscissa. It shows that 95% of all buildings larger than 50 m^2 and 90% of all buildings larger than 30 m^2 could be detected, whereas 96% of the detected building regions larger than 120 m^2 and 89% of all detected building regions were correct. The number of small buildings in

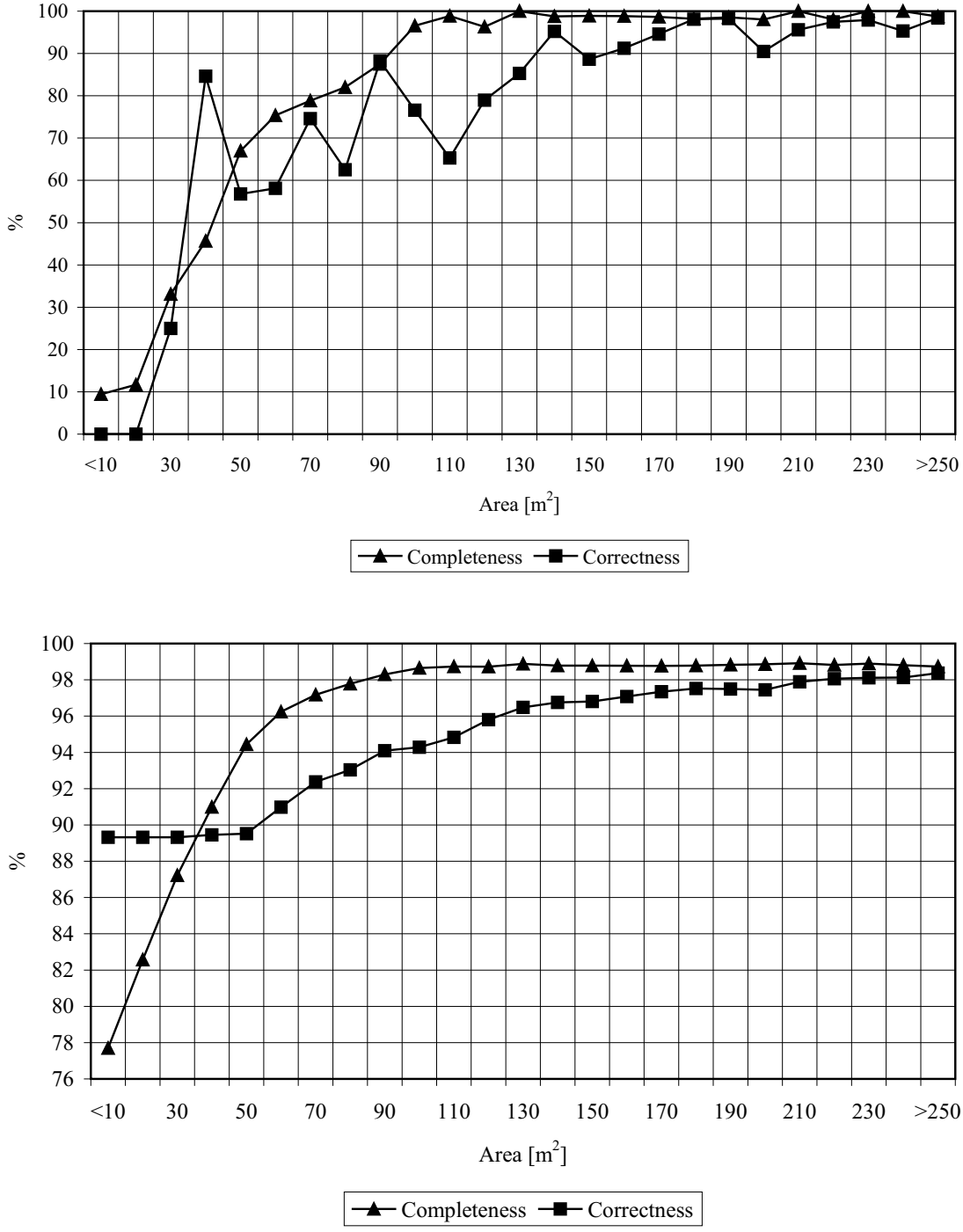


Figure 4.9: Above: completeness and correctness of buildings depending on the building size (class width: 10 m^2). Below: cumulative completeness and correctness of buildings larger than the size shown in the abscissa.

the reference data set was relatively large: 570 buildings or 24% covered an area smaller than 50 m^2, 296 or 12% were smaller than 30 m^2. On the other hand, only 30 buildings or 1.5% of the detected regions were smaller than 50 m^2. Based on figures 4.8 and 4.9, it is clear that the main buildings (larger than 90 m^2) were detected reliably by our method. The majority of the buildings between 50 and 90 m^2 could also be detected, whereas buildings smaller than 30 m^2 (mostly garden sheds or garages) could not usually be detected.

A comparison of these results with those presented in literature is difficult for two reasons: first, the test data are not identical, so that any comparison is somewhat uncertain, and second, the methodologies used for evaluation are not standardised and thus are often different. For instance, most authors do not give detection rates depending on the building size, and others only give per-pixel evaluations. In [Vögtle and Steinle, 2003], buildings were extracted from LIDAR data of 1 m resolution in rural and urban test areas, with detection rates of 92.6% and 95.8%, respectively. No information is given about the sizes of the buildings in the rural area (that could be compared to ours), and we do not know how the reference data were collected. The authors also state that the classification accuracy decreases with the building size. In [Matikainen et al., 2003], building change detection was carried out using LIDAR data sampled at 0.6 m. The authors state that 90% of all building pixels in a reference map were correctly detected, whereas 80.3% of the detected building pixels were also building pixels in the reference map. These numbers correspond to our completeness and correctness numbers, and are of a similar size. The authors also give completeness and correctness numbers separately for buildings larger than 200 m^2 and smaller than that threshold. Their numbers confirm the trend that can be seen in figure 4.9: completeness and correctness are 91.1% and 84.1%, respectively, for buildings larger than 200 m^2. For buildings smaller than 200 m^2, completeness and correctness are given by 42.1% and 34.9%, respectively. A minimum percentage of overlap of 70% is required for a building to be correctly identified [Matikainen et al., 2003]. We believe that the results in our tests demonstrate similar if not better results than these tests, and hence justify the approach taken for automatic building extraction in this research.

4.4 Conclusions and Future Work

We have presented a method for building detection from LIDAR data and multi-spectral images, and we have shown its applicability in a test site of heterogeneous building shapes. The method is based on a hierarchical technique for DTM generation and on the application of the Dempster-Shafer theory for classification. The results achieved were very satisfactory. 95% of the buildings larger than 50 m^2 can be detected, whereas about 89% of the detected buildings are correct. The detection rates decrease considerably with the building size: building structures smaller than 30 m^2 could generally not be detected. As a general trend, there were more false positives than false negatives. The quality of our results is comparable to those achieved by other research groups. Although this comparison is not conclusive because no common data set was used and because the evaluation methods are not always comparable, it shows that the Dempster-Shafer theory is well-suited for building detection.

Future work will concentrate on the influence of the parameters of the probability mass function used in this study on the results. We are also interested in the relative contribution of the individual cues to the classification results, especially in improving the classification results by the multi-spectral data. Another topic of future research is the replacement of the heuristic model for the probability mass functions by an empirical one that could be derived from the original classification results in an iterative procedure. Moreover, an investigation of the classification accuracy in

dependence of the LIDAR resolution would be interesting. We expect the second Dempster-Shafer classification stage to give better results with data of a higher resolution. We also expect the classification to be better if real near-infrared data are used, which is still to be investigated. Finally, we do not want the results of building detection as presented in this chapter to be the end of this research. These results are a prerequisite to geometrical reconstruction of buildings from roof planes, which we should be able to detect from building regions by fusing aerial imagery and LIDAR data. We also want to use the results of the first Dempster-Shafer classification and the final results of building extraction to improve the quality of the DTM by eliminating points on the building roofs before applying robust linear prediction as described in [Briese et al., 2002].

Acknowledgements

This work was supported by the Australian Research Council (ARC) under Discovery Project DP0344678 and Linkage Project LP0230563. The Fairfield data set was provided by AAM Geoscan, Kangaroo Point, QLD 4169, Australia (www.aamgeoscan.com.au).

Chapter 5

Building Detection by Fusion of Airborne Laserscanner Data and Multi-spectral Images: Performance Evaluation and Sensitivity Analysis

5.1 Introduction

Point clouds generated by airborne laser scanning (ALS) are well suited for the automatic detection of buildings. Building detection essentially requires a classification of the input data that separates points situated on buildings from those on other objects, especially trees. In order to accomplish this classification, cues such as the height of ALS points above the terrain or the roughness of the surface described by the ALS points can be used. Additional information can be considered in order to overcome problems occurring with buildings which consist of roof planes that are small in relation to the ALS resolution. These include the height differences between the first and the last echoes of the laser pulse and multi-spectral images of the area. The normalised difference vegetation index (NDVI), derived from multi-spectral images, is well suited for classification in this context [Lu et al., 2006].

Various classification techniques have been applied for building detection, e.g., unsupervised classification [Haala and Brenner, 1999], rule-based classification [Rottensteiner and Briese, 2002], Bayesian networks [Brunn and Weidner, 1997], [Stassopoulou et al., 2000], and fuzzy logic [Vögtle and Steinle, 2003, Matikainen et al., 2003]. The probabilistic approaches among the cited ones face the difficulty of having to model all a priori probabilities, which is problematic if the assumption of a normal distribution of the data vectors is unrealistic, e.g. in built-up areas [Gorte, 1999]. The theory of Dempster-Shafer can help to overcome these problems, because its capability to handle incomplete information provides a tool to reduce the degree to which assumptions about the distribution of the data have to be made [Klein, 1999].

Achieving good results for building detection using an algorithm demonstrates that the method works for the particular application, but it is also important to know how the parameters used in the algorithm are selected. For instance, if parameter selection is based on trial-and-error only, the reproducibility of the results for another data set is questionable. We consider the evaluation of algorithms to be important in order to make different approaches comparable. However, most

authors give detection rates and false alarm rates for the detected buildings, but fail to give a more thorough evaluation of their algorithms. Questions remaining unanswered in this context are related to the dependency of the results on scene and sensor characteristics, the availability of different input data sets, or on the appropriate selection of sensor models and the tuning of the model parameters. In this chapter, we especially want to deal with the evaluation of a method of automatic building detection. We have given an extensive overview of classification methods for building detection in [Rottensteiner et al., 2005b]. Therefore this chapter will commence with a review of papers dealing with the evaluation of building detection methods.

For an evaluation of automatic feature extraction using a reference data set, two numbers of interest are the completeness and the correctness of the results [Heipke et al., 1997]:

$$Completeness = \frac{TP}{TP + FN} \tag{5.1}$$

$$Correctness = \frac{TP}{TP + FP} \tag{5.2}$$

In equations 5.1 and 5.2, TP denotes the number of true positives, i.e., the number of entities found to be available in both data sets. FN is the number of false negatives, i.e., the number of entities in the reference data set that were not detected automatically, and FP is the number of false positives, i.e., the number of entities that were detected, but do not correspond to an entity in the reference data set.

[Vögtle and Steinle, 2003] evaluated their method of building detection using two test data sets of 1 m resolution and achieved detection rates of 93% and 96%, respectively. The authors state that the classification accuracy decreases with the building size, without quantifying this effect. [Matikainen et al., 2003] used ALS data for building change detection. Their method detected 90% of all building pixels in a reference map, with a false-alarm rate of 15%. On a per-building basis, completeness and correctness are 91% and 84%, respectively, for buildings larger than 200 m^2. The respective values for buildings between 0 and 200 m^2 are 42.1% and 34.9%. A minimum percentage overlap of 70% between a detected building and a building in the reference data set is required for the building to be classified as a true positive.

[Vosselman et al., 2004] first separate bare earth ALS points from other points and then further classify the other points according to whether they belong to buildings or trees. They apply their classification to the original ALS point clouds. Their results for points on buildings correspond to a completeness of 85% and a correctness of 92%. In their conclusions they state that using additional colour information increased the classification accuracy for buildings by 3%.

In our previous work, we have presented a method for fusing first and last pulse ALS and multi-spectral image data based on the theory of Dempster-Shafer. Completeness and correctness were evaluated for a test site in Australia [Rottensteiner et al., 2005b]. The main goals of this chapter are to present that method in its revised form and to thoroughly evaluate that method using two test sites of different land cover and sensor characteristics. From that evaluation, we want to assess the applicability of our method to different scenes and to data from sensors having different characteristics by finding answers to questions that are not commonly investigated by other authors:

- How realistic are the model assumptions about the properties of the sensor data?
- How can the control parameters be tuned?

- How sensitive are the results to the settings of these control parameters?
- How do the individual cues used in data fusion contribute to the quality of the classification results?
- How do the classification results deteriorate with decreasing sensor resolution?

We start with a description of the two test data sets in section 5.2. In section 5.3, we will give an outline of our previous work, describing the original algorithm for building detection. In section 5.4, we will present how that algorithm has been improved. In this context, the statistical models used for classification are evaluated and the rules for parameter tuning will be presented. This is followed by an evaluation of the building detection results in section 5.5. We not only present results obtained for standard parameter settings, but also include a sensitivity analysis with respect to the input parameters and the resolution of the input data, and we assess the impact of the individual classification cues. Section 5.6 will give the conclusions.

5.2 The Test Data Sets

We have used two test data sets. The first data set, captured over Fairfield (Australia) using an Optech ALTM 3025 laser scanner, was also used in the earlier study. The second data set was captured over Memmingen (Germany) with a TopoSys scanner. Both cover an area of 2×2 km^2, and both contain the first and the last echoes of the laser beam. The characteristics of the two test areas are quite different. Fairfield covers a suburban area with low density of development in the southwest half of the scene, whereas the northeast part is dominated by large industrial buildings. The trees are mostly evergreen. Memmingen features a densely developed historic centre in the north of the scene and industrial as well as suburban and rural areas in the remainder. The Memmingen data set was captured at a time when the trees had minimum foliage, so that a much larger proportion of last pulse points were reflected by the ground than in the Fairfield data set. The multi-spectral information was also quite different for the two data sets. For Fairfield, an RGB orthophoto with a resolution of 0.15 m was available (figure 5.1). We created a "pseudo-NDVI-image" at a resolution of 1 m, using the red band from the orthophoto and substituting the ALS intensity values for the infrared band. For Memmingen, geo-coded RGB (figure 5.1) and CIR (colour infrared) images with a resolution of 0.5 m were available. The infrared bands of the two data sets thus correspond to different wavelengths, namely 1047 μm for Fairfield and 770-890 μm for Memmingen.

The two sensors involved have a different scanning pattern. For the Optech ALTM 3025 used for Fairfield, the scanning pattern was fairly regular with an average point separation of about 1.2 m. It was only at the edges of the scanning swaths that the point distribution was not homogeneous. The TopoSys scanner delivers very high point densities in the flight direction of 0.2 m, but a relatively coarse distribution of points across the flight direction of 1.2 m. From the original ALS point clouds, raster-based Digital Surface Models (DSMs) were interpolated for both the first and the last pulse echoes by linear prediction using a small degree of smoothing [Rottensteiner and Briese, 2002]. The width Δ of the DSM grids was chosen to be Δ = 1 m for Fairfield and Δ = 0.75 m for Memmingen.

Reference data were captured by digitising buildings and trees in the orthophotos. We chose to digitize all structures recognisable as buildings or trees independent of their size. Neighbouring buildings that are joined but obviously separate entities were digitized as separate polygons. This

Figure 5.1: Intensity bands of the orthophotos for Fairfield (left) and Memmingen (right). Both cover an area of 2×2 km^2.

resulted in 2424 building polygons in Fairfield[1] and 2046 in Memmingen. Larger areas covered by trees were digitised as one polygon. No information on single trees was captured. In Fairfield, the orthophoto and the ALS data correspond to different epochs. We thus had to exclude 49 building polygons that were only available in one data set.

5.3 The Original Algorithm for Building Detection

The input to our method comprises four data sets that have to be generated from the raw data by pre-processing: the two DSM grids corresponding to the first and the last pulse data; a Digital Terrain Model (DTM); and the NDVI. The DTM can be derived from the last pulse DSM by hierarchic morphologic filtering [Rottensteiner et al., 2005b] or by robust linear prediction [Rottensteiner and Briese, 2002]. The classification itself is based on the theory of Dempster-Shafer for data fusion [Klein, 1999].

In Dempster-Shafer fusion, the output of a set of "sensors" is used for a classification process in which n disjunct classes $C_j \in \theta$ are to be discerned. Denoting the power set of θ by 2^θ, a *probability mass* fulfilling certain constraints has to be assigned to every class $A \in 2^\theta$ (i.e., also to any combination of the original classes) by each sensor. The probability masses from the individual sensors can be combined, and from these combined probability masses, two parameters can be computed for each class: the *Support* of a class is the sum of all masses assigned to that class, and the *Plausibility* sums up all probability masses not assigned to the complementary class of A. The accepted hypothesis $C_a \in \theta$ is determined according to a decision rule. The Dempster-Shafer theory also provides a measure for

[1]This number differs from the one given in chapter 4 (2385). The difference is caused by buildings that were originally missed in the reference data set.

the *Conflict* in the sensor data [Klein, 1999].

Building detection starts with a Dempster-Shafer fusion process carried out for each pixel of the DSM grid in order to achieve a classification of the input data into one of four classes: buildings (B), trees (T), grass land (G), and bare soil (S), thus $\theta = \{B, T, G, S\}$. In [Rottensteiner et al., 2005b], we presented a model for the distribution of the evidence from each sensor used for classification to the four classes, assuming that each sensor i can separate two complementary subsets subsets U_{Ci} and $\overline{U}_{Ci}$ of θ. The probability mass $P_i(x_i)$ assigned to U_{Ci} by the sensor i depending on the sensor output x_i is assumed to be equal to a constant P_l for $x_i < x_l$. For $x_i > x_u$, it is assumed to be equal to another constant P_u, with $0 \leq P_l < P_u \leq 1$. Between x_l and x_u, the probability mass is modelled by a cubic parabola with horizontal tangents at $x_i = x_l$ and $x_i = x_u$:

$$P_i(x) = \begin{cases} P_l & \forall x | x \leq x_l \\ P_l + (P_u - P_l) \cdot \left[3 \cdot \left(\frac{x - x_l}{x_u - x_l} \right)^2 - 2 \cdot \left(\frac{x - x_l}{x_u - x_l} \right)^3 \right] & \forall x | (x > x_l) \wedge (x < x_u) \\ P_u & \forall x | x \geq x_u \end{cases} \tag{5.3}$$

The probability mass $[1 - P_i(x_i)]$ will be assigned to $\overline{U}_{Ci}$. It is one of the advantages of the Dempster-Shafer theory that no other assumptions about the distributions of the data with respect to the classes are required. The combined probability masses are computed for each pixel, and the pixel is assigned to the class of maximum support or maximum plausibility. We use five "sensors" for the original classification:

- The height differences ΔH between the DSM and the DTM help to distinguish elevated objects from the ground, thus $U_{C\Delta H} = B \cup T$.
- The strength of surface roughness R is defined as the texture strength of polymorphic feature extraction [Förstner, 1994] applied to the first derivatives of the DSM. It corresponds to the smoothed squared sum of the second derivatives of the DSM within a small window. Large variations of the first derivatives of the DSM are typical for trees, thus $U_{CR} = T$.
- The directedness D of surface roughness, i.e. the texture directedness of polymorphic feature extraction [Förstner, 1994], is another indicator for trees, i.e. $U_{CD} = T$. However, D is only to be used if R differs significantly from 0.
- The height differences ΔH_{FL} between the first and the last pulse DSMs also distinguish trees from other classes; $U_{C\Delta HFL} = T$.
- The $NDVI$ is an indicator for vegetation, thus $U_{CNDVI} = T \cup G$. The NDVI is used instead of the complete multispectral information because it is directly linked to our actual problem, i.e. the discrimination of buildings and trees, and its properties for solving that problem are easily modelled. This is not the case for the individual bands of the multi-spectral images, because the distributions of the grey levels of the individual bands depend on surface properties. This means for instance that there are different clusters in feature space for buildings having different roof colours.

In [Rottensteiner et al., 2005b], we gave values for (x_l, x_u) for all of these five sensors without any further discussion, assuming they were generally applicable. We will see in section 5.4.1 that

an evaluation of the statistical model described by equation 5.3 showed more precisely how to select these parameters. This evaluation led to a reformulation of some of the rules for assigning the probability masses.

Initial building regions are determined as connected components of "building pixels". A second classification based on the Dempster-Shafer theory is applied to these initial building regions, using four cues representing average values for each building region, namely the average height differences ΔH_a between the DSM and the DTM, the percentages H and P of pixels classified as "homogeneous" and "point-like", respectively, in polymorphic feature extraction, and the average $NDVI_a$. The reader is referred to [Rottensteiner et al., 2005b] for further details of the method.

5.4 Evaluation of the Statistical Model and Improvements to the Algorithm

In this section, we will describe changes to the models for the probability masses in the initial land cover classification that are essential so that the models are more realistic, and a new post-classification technique. The section also contains a discussion on the selection of the model parameters and an empirical validation of our theoretical models.

5.4.1 Initial Land Cover Classification

5.4.1.1 Height Differences ΔH Between DSM and DTM

ΔH is used in the same way as described in section 5.3. The last pulse DSM should be used to optimise the classification accuracy for buildings. Figure 5.2 compares $P_{\Delta H}(\Delta H)$ derived from the ground truth to the theoretical model with $(P_l, P_u) = (5\%, 95\%)$ and $(x_l, x_u) = (0\ m, 4\ m)$. The theoretical model fits quite well to the data of both Memmingen and Fairfield. The differences for small values of ΔH correspond to pixels within the tree polygons having a last pulse return from the terrain. The parameters (x_l, x_u) should be selected so that $x_C = (x_l + x_u)/2$ is larger than the minimum height expected for a building. We consider the parameter values given above ($x_C = 2\ m$)

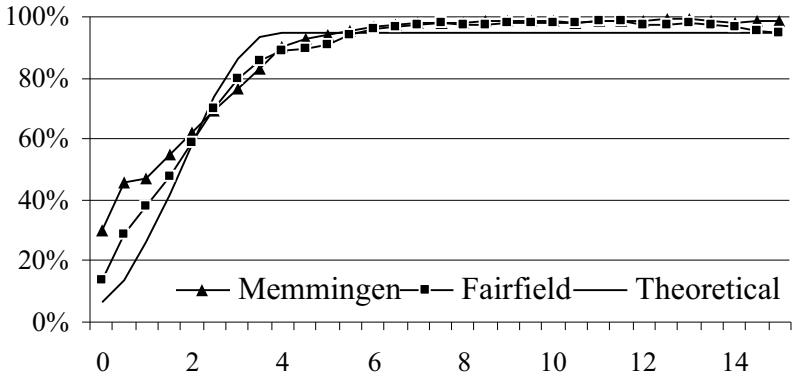


Figure 5.2: $P_{\Delta H}(\Delta H)$ for $B \cup T$ from ground truth compared to the theoretical model. Abscissa: $\Delta\ H$ [m].

to be reasonable and meaningful, and thus we think that these parameters only require special tuning if the buildings in a scene have different height characteristics.

5.4.1.2 Strength of Surface Roughness

As pointed out in section 5.3, R was originally chosen to correspond to the smoothed squared sum of the second derivatives of the DSM within a small window. In [Rottensteiner et al., 2005b], we selected $(x_l, x_u) = [k_l \cdot median(R), k_u \cdot median(R)]$, using $(k_l, k_u) = (2, 15)$. By making the two thresholds dependent on the median of R, we expected them to be adaptive to the scene content; the actual parameters to be chosen were k_l and k_u. However, a comparison of the theoretical model for the probability masses to ground truth showed considerable deviations, and the shapes of the empirical curves were heavily dependent on the scene contents. It turned out to be impossible to find values for the parameters of the model described by equation 5.3 that are generally applicable, and there was no easy way to estimate appropriate values for the parameters k_l and k_u. This situation was improved by a re-parameterisation of R. Rather than using R, we characterise surface roughness by the percentage $R_P(R)$ of pixels for which the surface roughness is smaller than R. $R_P(R)$ is limited to the interval $[0 \ldots 100\%]$.

We assign a probability mass $P_R = P_R(R_P)$ to class T, and $(1 - P_R)$ to $B \cup G \cup S$, neglecting that large values of R might also occur at the borders of buildings. Assuming that the trees correspond to the areas of maximum surface roughness, we can use an estimate for the percentage P_T of the scene covered by trees to derive the values of the parameters of the model in equation 5.3. The P_T percent "roughest" pixels should correspond to the trees. This should be reflected by a probability mass $P_R > 0.5$ for these pixels, thus x_l can be determined so that $P_R(100\% - P_T) = 0.5$. Using $P_u = 1 - P_l$ and $x_u = 100\%$ yields:

$$x_l = 2 \cdot 100\% - x_u - 2 \cdot P_T = 100\% - 2 \cdot P_T \tag{5.4}$$

In this way, we succeed in replacing the selection of the two parameters k_l and k_u that do not have a physical meaning by the "meaningful" parameter P_T that can be estimated in a comparably simple way on the basis of a visual inspection of the data.

On the basis of estimates for P_T, this results in $x_l = 70\%$ for Fairfield and $x_l = 50\%$ for Memmingen. Figure 5.3 compares the distribution of $P_R(R_P)$ from ground truth with the theoretical values thus obtained. It shows that the model fits quite well for the last pulse data in Fairfield, with a larger deviation for first pulse data which is mainly caused by powerlines. With Memmingen, the model fits quite well to the first pulse data, but not so well to the last pulse data. This is caused by the high penetration rate of the laser in this data set. There are many laser strikes on the ground in the forested areas, which means that the DSM almost corresponds to the (smooth) DTM in these areas. As a consequence, and in contrast to the application of the algorithm described in section 5.3, we let the user decide whether to use first or last pulse data for computing surface roughness, depending on which data set best represents the roughness of the DSM for vegetation. In Fairfield, we use last pulse data, in Memmingen, first pulse. The most important parameter is the estimate for for the percentage P_T of the scene that is covered by trees.

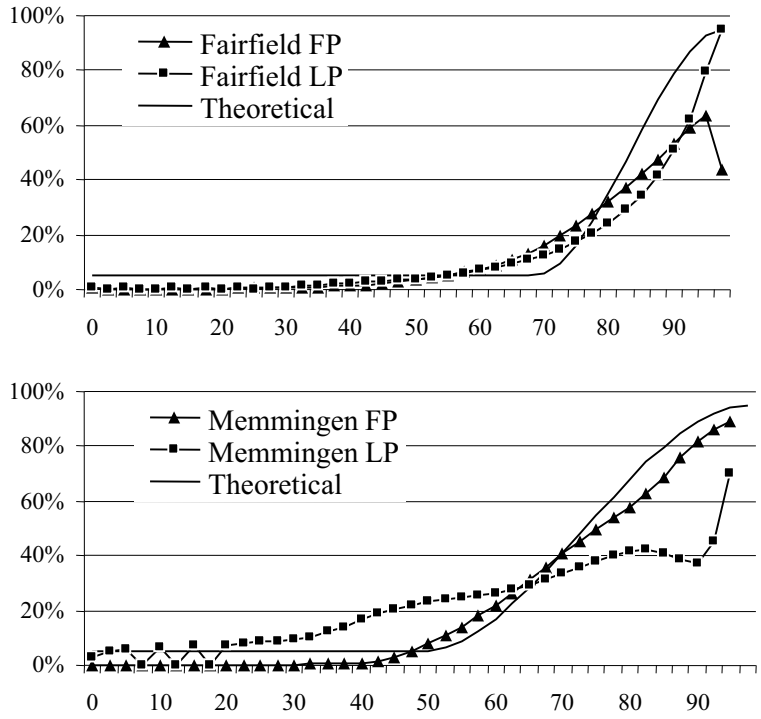


Figure 5.3: $P_R(R_P)$ for T from ground truth compared to the theoretical model for Fairfield (above) and Memmingen (below). LP/FP: Last / First pulse data. Abscissa: R_P [%].

5.4.1.3 Directedness D of Surface Roughness

The directedness D of surface roughness is used in the same way as described in section 5.3. The considerations with respect to using first or last pulse data described for the strength R of surface roughness in section 5.4.1.2 also hold true for D. In order to decide whether R is significant, we have to compare it to a threshold R_{min}. Again, the re-parameterisation of R as described in section 5.4.1.2 helped to find a better way of selecting that threshold than the one described in the earlier study

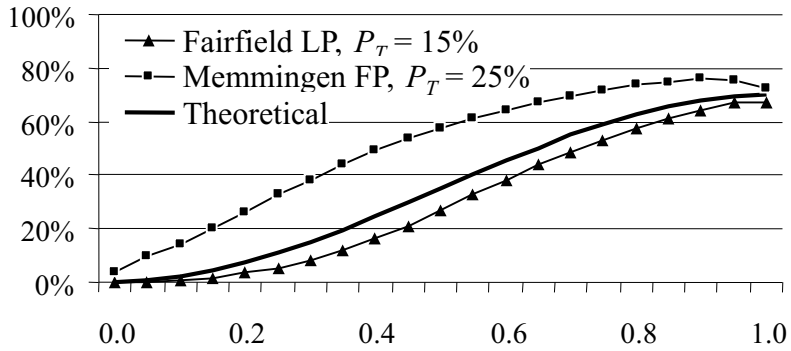


Figure 5.4: $P_D(R > R_{min}, D)$ for T from ground truth compared to the theoretical model. Only DSM Pixels with $R > R_{min}$ are considered. R_{min} is derived from P_T. Abscissa: D.

[Rottensteiner et al., 2005b]. We determine R_{min} so that for P_T percent of the data $R > R_{min}$; thus, D will be considered for the P_T "roughest" pixels. We select $(P_l, P_u) = (10\%, 70\%)$ and $(x_l, x_u) = (0, 1)$. The lower value for P_u compensates for the fact that large values for D occur with buildings having many small roof planes. Figure 5.4 shows the distribution of D and the theoretical values, using the same values for P_T as for R. It fits quite well for the last pulse data in Fairfield. There are some deviations in the Memmingen data set, but the trend is correct. Again, the most important parameter is the estimate for the percentage P_T of the scene covered by trees.

5.4.1.4 Height Differences ΔH_{FL} Between First and Last Pulse

ΔH_{FL} is used in a similar way as described in section 5.3. We found that, whereas a large value of ΔH_{FL} certainly does give support to the hypothesis that there is a tree, a small value of ΔH_{FL} does not necessarily mean that there is no tree. In order to model this behaviour, we assign the probability mass $(1 - P_{FL})$ according to the model described in section 5.3 to θ rather than to the complementary class of T (i.e., $B \cup G \cup S$). We consider the values $(P_l, P_u) = (5\%, 95\%)$ and $(x_l, x_u) = (0\,m, 4\,m)$ to be generally applicable. Figure 5.5 compares $P_{FL}(\Delta H_{FL})$ derived from ground truth to the theoretical model. The empirical curves are not quite as steep as the theoretical one, which is caused by large height differences between first and last pulses at building outlines. In Fairfield, the upper limit for the probability mass seems to be 80% rather than 95%, which is caused by large values of ΔH_{FL} at powerlines. This mostly results in a misclassification of grass-land or bare soil as trees and thus is not critical for our goal of detecting buildings.

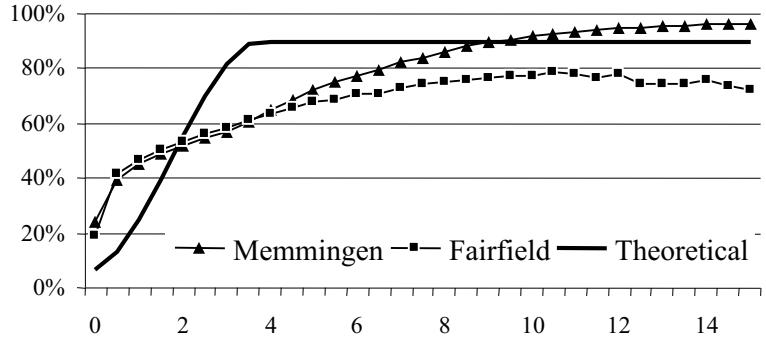


Figure 5.5: $P_{FL}(\Delta H_{FL})$ for T from ground truth compared to the theoretical model. Abscissa: $\Delta\ H_{FL}$ [m].

5.4.1.5 NDVI

As described in section 5.3, the NDVI is an indicator for vegetation, thus for classes T and G. However, the model for the probability masses described in section 5.3 had to be improved to consider the uncertainty of the NDVI in shadow areas.

The amount of radiation arriving at a passive sensor depends on many factors. Some of them are characteristic for the sensor or the object, but there are also geometrical ones, namely the direction of the sun, the sensor viewing direction, and the normal vectors of the illuminated surfaces. A slope facing the sun will appear brighter than a slope pointing away from the sun. In remote

sensing, this is called "topographic effect", and various methods have been proposed in the past to eliminate it, e.g. [Hejmanowska, 1998]. In this work, rather than modelling the physical reality of the light's interaction with the surface e.g. by a Bidirectional Reflectance Distribution Function [Hejmanowska, 1998], which requires parameters such as the directional vector of the sun to be known, we will use the observation that the uncertainty of the NDVI is larger in dark areas than in bright ones in order to the reduce the classification bias introduced by the topographic and shadow effects.

The NDVI is defined as the ratio between the difference and the sum of the infrared band IR and the red band Rd, thus $NDVI = 100\% \cdot (IR - Rd)/(IR + Rd)$. Applying the laws of error propagation to this equation, the standard deviation σ_{NDVI} of the NDVI can be computed from the standard deviations σ_{Rd} and σ_{IR} of Rd and IR:

$$\sigma_{NDVI} = \frac{2 \cdot \sqrt{Rd^2 \cdot \sigma_{IR}^2 + IR^2 \cdot \sigma_{Rd}^2}}{(IR + Rd)^2} \cdot 100\% \tag{5.5}$$

The standard deviations σ_{Rd} and σ_{IR} are determined by analysing the first derivatives of IR and Rd [Förstner, 1994]. Figure 5.6 shows the intensity of the CIR image, the NDVI, and σ_{NDVI} for a part of the Memmingen data set. For buildings with a ridge in east-west direction, the NDVI is 25% in the shady roof planes and -10% in the sunny areas, with standard deviations of about $\pm$10% and $\pm$2%, respectively. The NDVI suggests a strong support for classifying the roof planes in the shade as vegetation, but this support is mitigated by the fact that it is also relatively uncertain. In order to compensate for the bias of the NDVI in shadow areas, probability masses for the NDVI are modulated depending on σ_{NDVI}. For $\sigma_{NDVI} > 25\%$, we assign a probability mass of 1.0 to θ, i.e., the NDVI will not contribute to the classification. If $\sigma_{NDVI} \leq 25\%$, $P_{N\theta} = 2 \cdot \sigma_{NDVI}$ is assigned to θ. Using $P_N^0 = P_N(NDVI)$ according to equation 5.3, we assign $P_{NDVI} = (1 - P_{N\theta}) \cdot P_N^0$ to class $T \cup G$ and $P_{Ninv} = (1 - P_{N\theta}) \cdot (1 - P_N^0)$ to $B \cup S$. Note that the problems with the uncertainty of the NDVI did not show up with the Fairfield data set, where IR was taken from the ALS intensities to compute a "pseudo-NDVI" (cf. section 5.2): ALS is an active sensing technique, and there are no cast shadows in the ALS intensity. It turns out that this adaptation of the model for the probability masses of the NDVI does not necessarily improve the detection rates of the buildings, but it does improve the shapes of the detected buildings in the shadow areas.

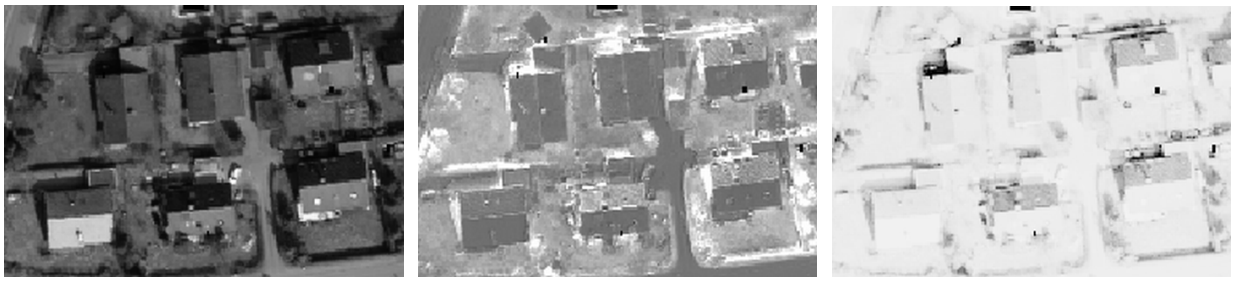


Figure 5.6: Intensity of the CIR image (left), NDVI (centre) and σ_{NDVI} (right) for a part of the Memmingen data set. In the σ_{NDVI} image, white areas correspond to low σ_{NDVI}.

As the classes G and S are very broadly defined, it is impossible to digitize ground truth data for these classes in a meaningful way. We thus do not have ground truth for the classes G and S, so that we cannot check our model for the NDVI. As the NDVI depends on the lighting conditions, on

the predominant type of vegetation, and on the wavelength at which the infrared band was recorded, we think that the parameters for the model have to be determined in a training phase. Here we choose $(P_l, P_u) = (10\%, 90\%)$ for both data sets. Further, we choose x_l and x_u to be (-30%, 30%) for Fairfield and (-10%, 30%) for Memmingen.

5.4.2 Post-Classification and Generation of Initial Building Regions

A new rule-based technique for post-classification has been developed in order to eliminate single building pixels and to compensate for classification errors at the building outlines. Two steps are carried out in an iterative way until a maximum number of iterations is reached. First, we check pixels where the classification is uncertain due to contradicting evidence as indicated by a conflict $K > 50\%$; the conflict measure is one of the results of the Dempster-Shafer analysis (cf. section 5.3). For these pixels, we compute a histogram of the classes in a 5×5 neighbourhood. If the class $C_{5\times5}$ having the maximum number of occurrences is identical to the class C_2 achieving the second highest score according to the decision rule in the Dempster-Shafer fusion, the classification of the pixel is changed to C_2. In the second step, we determine the class $C_{3\times3}$ having a maximum number of occurrences in a 3×3 neighbourhood of each pixel. The classification is changed if $C_{3\times3}$ is identical to C_2 or if all pixels except the central one are assigned to $C_{3\times3}$. A binary building image is created from all pixels classified as "building" after the final iteration. Morphologic opening is used to eliminate small areas of building pixels and to separate weakly connected buildings. Then, the initial building regions are determined by a connected component analysis of the binary building image.

5.4.3 Region-based Classification

The second classification step is carried out in a similar way described in section 5.3, with three adaptations: in computing the average NDVI ($NDVI_a$), the individual NDVI values are weighted by $1/\sigma^2_{NDVI}$, the percentage H of pixels classified as "homogeneous" in polymorphic feature extraction is no longer used, and the threshold separating "homogeneous" from "inhomogeneous" areas is chosen so that the percentage of the pixels that will be classified as "inhomogeneous" is identical to the percentage P_T of the scene covered by trees. Whereas for $NDVI_a$ and for the average building height ΔH_a similar parameter settings as for the NDVI and for ΔH can be chosen, tuning the parameters for the percentage P of pixels classified as "point-like" in polymorphic feature extraction (cf. section 5.3) is more difficult, as P also depends on P_T and on the average size of a roof plane in relation to the ALS resolution. It turns out that the improved model for the probability masses described in section 5.4.1 and the new post-classification technique described in section 5.4.2 have the effect that this second classification step has almost no influence on the classification results. For reasons of completeness, we give the values used for the statistical model given by equation 5.3 in this study: for $NDVI_a$ we use the same parameter settings as for the NDVI, but without the modulation by the standard deviation σ_{NDVI}. For ΔH_a, the values for $(x_l, x_u) = (1\ m, 3\ m)$ are chosen to be a bit tighter than for ΔH . For P we selected $(x_l, x_u) = (25\%, 75\%)$, based on a training phase.

5.5 Results and Evaluation

5.5.1 Methodology

The methodology for evaluation is based on the technique for comparing the classification results with a reference data set described in [Rottensteiner et al., 2005b], namely on a comparison of two label images: the "automatic label image", i.e. the output of the building detection algorithm, and the "reference label image" that is generated by rasterizing the reference polygons. We are interested in determining completeness and correctness according to equations 5.1 and 5.2 for two types of entities. Firstly, the analysis is done on a per-pixel level and secondly on a per-building level, showing how many buildings could be detected and how many of the buildings detected automatically did actually correspond to buildings. In this chapter we determine TP, FN, and FP (cf. equations 5.1 and 5.2) by the same way as described in detail in [Rottensteiner et al., 2005b], especially with respect to handling the problem of multiple overlaps for the per-building analysis.

In section 5.5.2, we present the results achieved using the standard parameter settings described in section 5.4. In section 5.5.3, we will assess the sensitivity of the method to the parameter settings in the first (and most important) classification stage by systematically changing these parameters and comparing the completeness and correctness values thus achieved. The influence of changing sensor resolution will be determined in the same way and is presented in section 5.5.4. Finally, in section 5.5.5 the contributions of the individual classification cues will be evaluated by varying the input data and comparing the completeness and correctness values. The comparison of different variants will be carried out only on a per-building basis, because it is more telling than the per-pixel comparison.

5.5.2 Classification Results Using the Standard Parameters

Figure 5.7 shows the results of the initial Dempster-Shafer classification after post-processing, using the modified approach as described in section 5.4. In Fairfield, 2057 building regions are detected after region-based classification. For Memmingen, the number of buildings is 2102. Step edges at the building boundaries are often classified as trees, an effect that is reduced but not completely eliminated by post-classification.

In Fairfield, 87% of the building pixels were detected. In Memmingen, this number was 91%. The missed buildings were mostly small residential buildings having roofs consisting of many small faces, or they were too small to be detected given the resolution of the ALS data. Correctness was somewhat better than completeness (91% in Fairfield, 92% in Memmingen). False positives mostly occur at bridges, at small terrain structures not covered by vegetation, and at container parks. These numbers have to be interpreted with caution because they are affected by errors in the reference data. The orthophotos were generated using a DTM, so that buildings are shifted away from the nadir point of the camera. In Fairfield, this shift can be up to 5 m.

Figure 5.8 shows the completeness and correctness for buildings depending on the building size. It clearly shows how these numbers decrease with decreasing area covered by the buildings. In the first row, completeness and correctness are given for buildings of the size shown in the abscissa. In Fairfield, buildings larger than 110 m^2 can be detected reliably, with both completeness and correctness being larger than 95%. Our algorithm could detect more than 80% of the buildings with an area between 80 m^2 and 110 m^2, and the majority of buildings with an area between 40 m^2 and 80 m^2

Figure 5.7: Results of the Dempster-Shafer classification after post-processing for Fairfield (left) and Memmingen (right). White: bare soil. Light grey: grass land. Dark grey: trees. Black: buildings.

could still be detected. Buildings smaller than 30 m^2 were not detectable. The results achieved for Fairfield are somewhat better than those described in our earlier study [Rottensteiner et al., 2005b]; the improvement might be due to the improved model for the surface roughness. With Memmingen, a greater percentage of smaller buildings could be detected, at the cost of a somewhat higher ratio of false positives with an area of 30 m^2 to 50 m^2. For buildings larger than 120 m^2 both completeness and correctness were larger than 95%. Looking at the cumulative completeness / correctness in the second row of figure 5.8, we can say that 95% of all buildings larger than 70 m^2 (Fairfield) and 50 m^2 (Memmingen) could be detected. In both data sets 95% of all detected buildings larger than 70 m^2 were correct. These results show that the algorithm performs consistently well for:

- Scenes having different characteristics as far as building shapes and vegetation type are concerned
- Data having a different original resolution
- Data having been captured by sensors of different characteristics with respect to both the way the data are captured and the scanning pattern.

5.5.3 Sensitivity Analysis

The results presented in the previous section were achieved using the standard parameter settings as explained in section 5.4. The most important control parameters are those used for the model of the probability masses in the initial classification, namely the values (x_l, x_u) for the models of the probability masses for the height differences ΔH between the DSM and the DTM, the height differences ΔH_{FL} between the first and the last pulse, and the NDVI, and the estimate P_T for the percentage

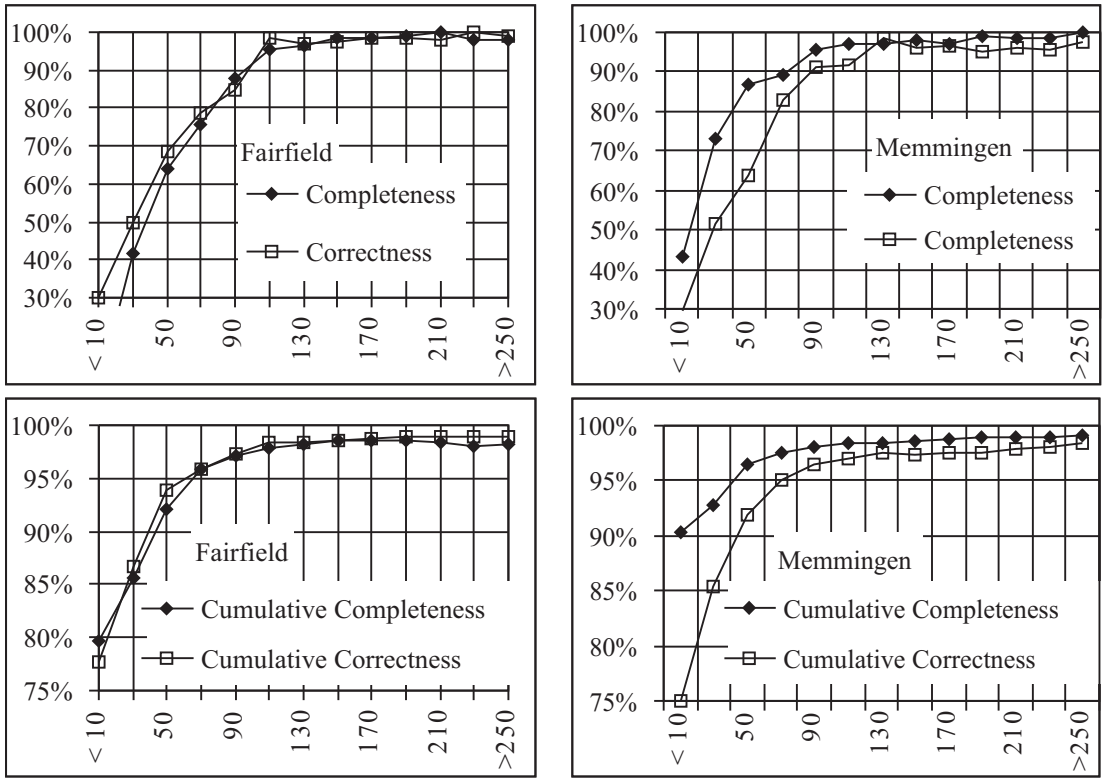


Figure 5.8: Completeness and correctness of the detection results as a function of the building size [m^2] for Fairfield (left) and Memmingen (right). First row: completeness and correctness for buildings of the size given in the abscissa. Second row: cumulative completeness and correctness computed for all buildings larger than the size given in the abscissa.

of trees in the scene, which is essential to derive the parameters for the probability masses of both surface roughness parameters (strength R and directedness D). In this section, we will evaluate the sensitivity of the results with respect to the setting of these parameters. Starting from the standard settings, each of these parameters will be systematically changed, and each time completeness and correctness will be computed, in order to assess how changes in the parameter settings affect the quality of the results. In all cases, the completeness and correctness curves will only be shown for buildings having an area smaller than 250 m^2 because for larger buildings, the method in general performs well in all cases.

5.5.3.1 Influence of the Parameters for the Model of the Probability Masses for ΔH and ΔH_{FL}

Starting from the standard values $(x_l, x_u) = (0\ m, 4\ m)$ for the height differences ΔH between the DSM and the DTM, eleven variants were computed by changing both x_l and x_u by full multiples of 1 m. It turned out that completeness and correctness were very similar for all variants having an identical centre point $x_C = (x_l + x_u)/2$. That is why only four characteristic variants for Fairfield and five for Memmingen are shown in figure 5.9. In both cases, the correctness seems to be less sensitive to variations of the model parameters for ΔH than the completeness, and the improvement in correctness achieved by choosing parameter settings different from the standard ones are outweighed

by the deterioration in completeness. In Fairfield, the influence of variations of the parameters for ΔH is much larger than in Memmingen. For the variant with $(x_l, x_u) = (1\ m, 5\ m)$, thus $x_C = 3\ m$, only about 60% of the buildings that cover an area of about 200 m^2 are detected. This means that in Fairfield, for the buildings covering an area of about 200 m^2, i.e. larger residential buildings, large building parts are lower than 3 m. In Memmingen a similar effect occurs for the variant with $(x_l, x_u) = (3\ m, 7\ m)$, thus $x_C = 5\ m$. This is consistent with the predominant building types. Changing x_C by 0.5 m in variant $(x_l, x_u) = (1\ m, 4\ m)$ has most effect on buildings smaller than 120 m^2 with both data sets. With variant $(x_l, x_u) = (1\ m, 5\ m)$, the deterioration in completeness is already in the range of 10% - 40% even for larger buildings.

We conclude that the parameters for the model for the probability masses for ΔH should be known relatively precisely (up to 0.5 m). They must be selected so that x_C is smaller than the heights of the lowest buildings that should be detected in the data. The standard parameters perform well with both data sets and thus seem to be applicable to a variety of scene characteristics, but scenes in city centres containing only high-rise buildings might favour other parameter settings.

For ΔH_{FL}, the same variants as for ΔH were computed. It turns out that changing the parameters hardly had any effect. This might be caused by the fact that some laser scanners can only separate pulses if the range difference is larger than a certain threshold. For the Fairfield data set, this threshold is about 4.6 m, whereas for the scanner used in Memmingen it is not known. In the future, full waveform laser scanners might provide additional information for discriminating trees from buildings.

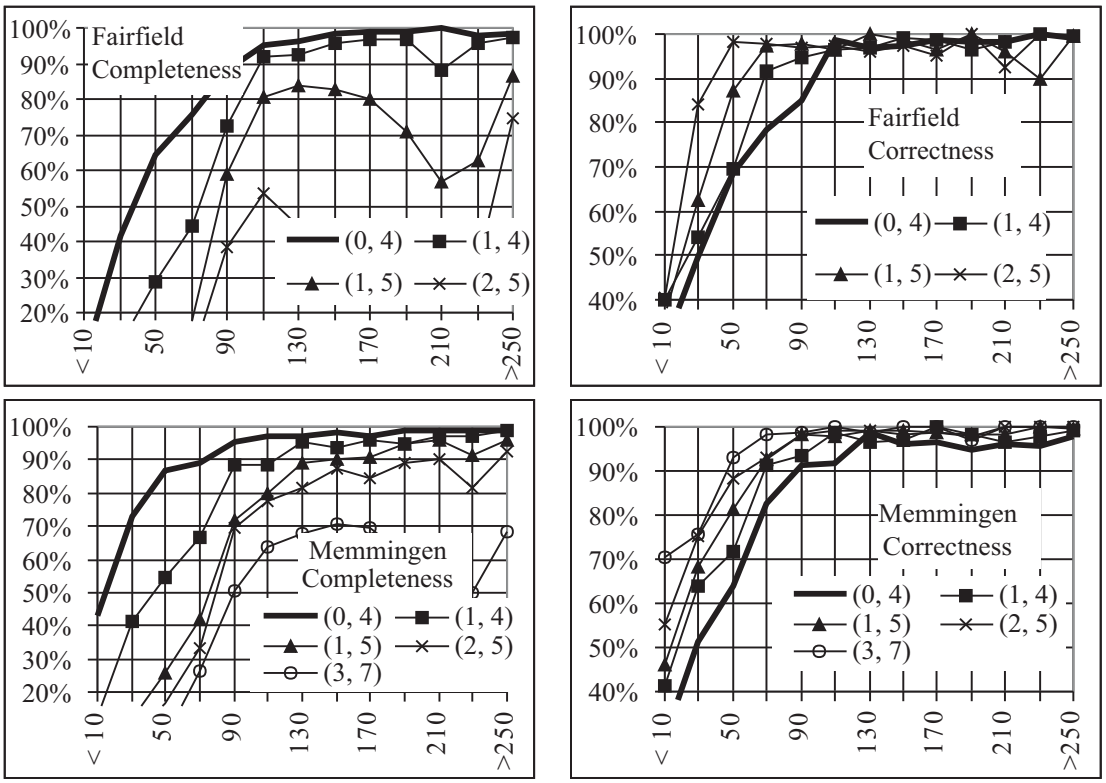


Figure 5.9: Completeness and correctness of the results for different settings of the parameters $(x_l,\ x_u)$ [m] for the model of the probability masses for ΔH as a function of the building size [m^2].

5.5.3.2 Influence of the Parameters for the Model of the Probability Masses for the NDVI

Starting from the standard values $(x_l, x_u) = (-30\%, 30\%)$ for Fairfield and $(x_l, x_u) = (-10\%, 30\%)$ for Memmingen, 12 different variants were computed by changing both x_l and x_u by multiples of 10%. Again it was observed that the results depended mainly on the selection of $x_C = (x_l + x_u)/2$. The results for five variants are shown in figure 5.10. The standard settings seem to be a good compromise, since good completeness values were achieved without too many false positives. In Fairfield, using $(x_l, x_u) = (-20\%, 40\%)$ would also have been an acceptable choice. The results are more sensitive to changes in the parameters of the model for the NDVI in Memmingen than in Fairfield. Changes of x_C of 10% reduce the completeness results for buildings smaller than 150 m^2 by 10% to 20%, without improving completeness accordingly. On the other hand, choosing x_C too large has less impact on the results than choosing it too low. With small buildings, it is more likely that a building is misclassified as a tree than vice versa: apparently, most trees have an NDVI substantially higher than x_C in the standard parameter settings. We conclude that the parameters for the model of the probability masses of the NDVI have to be known relatively well, and they have to be determined in a training phase.

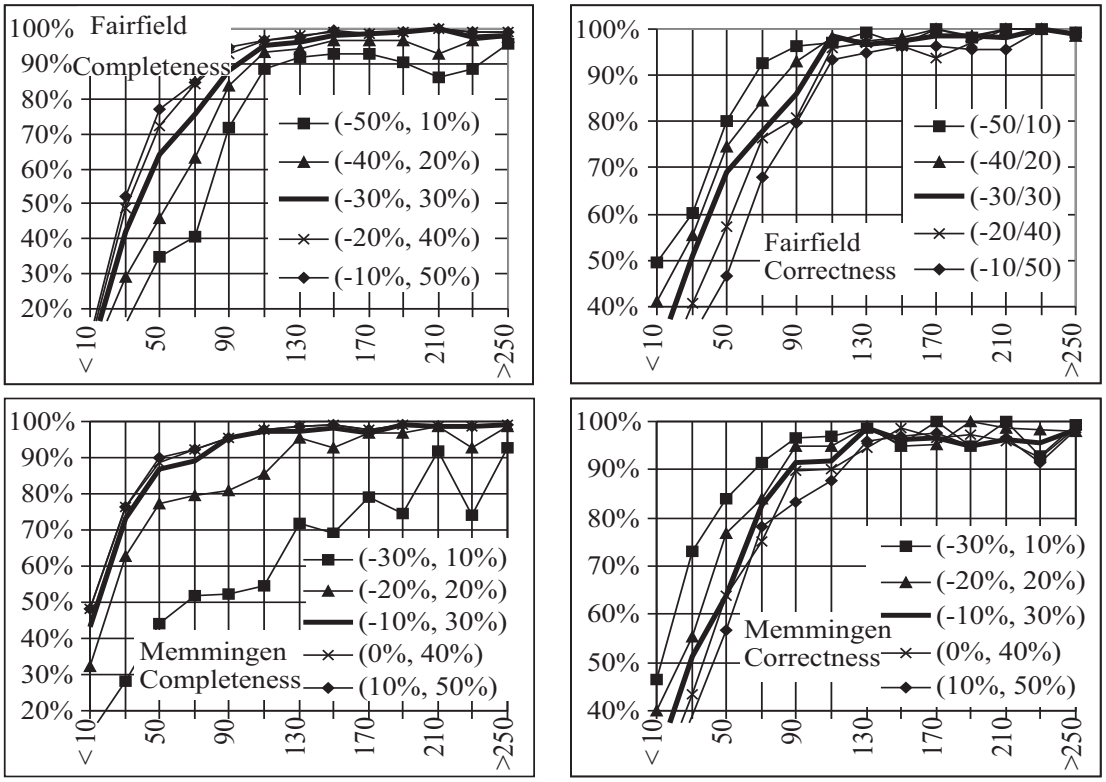


Figure 5.10: Completeness and correctness of the results for different settings of the parameters (x_l, x_u) [%] for the model of the probability masses for the NDVI as a function of building size [m^2].

5.5.3.3 Influence of the Estimate for the Percentage P_T of the Scene Covered by Trees

Starting from the standard values $P_T = 15\%$ for Fairfield and $P_T = 25\%$ for Memmingen, several variants were computed using values for P_T differing in multiples of 5% from the standard val-

ues, and completeness and correctness of the results were evaluated. The results are depicted in figure 5.11. Again, the standard parameters achieve a good trade-off between completeness and correctness. With Fairfield, a value for P_T between 15% and 20% might be considered optimal. Choosing P_T smaller than the estimated percentage of trees slightly improves the completeness of the results, but correctness deteriorates by a much bigger percentage. For completeness, choosing P_T too large seems to be more critical than choosing it too small. Figure 5.11 indicates that an over-estimation of P_T by only 10% causes a deterioration of completeness by about 10%-15% for buildings of about 100 m^2. We conclude that P_T should be known to within 5% in order to achieve optimal results, especially for smaller buildings, to be able to exploit the full potential of the method.

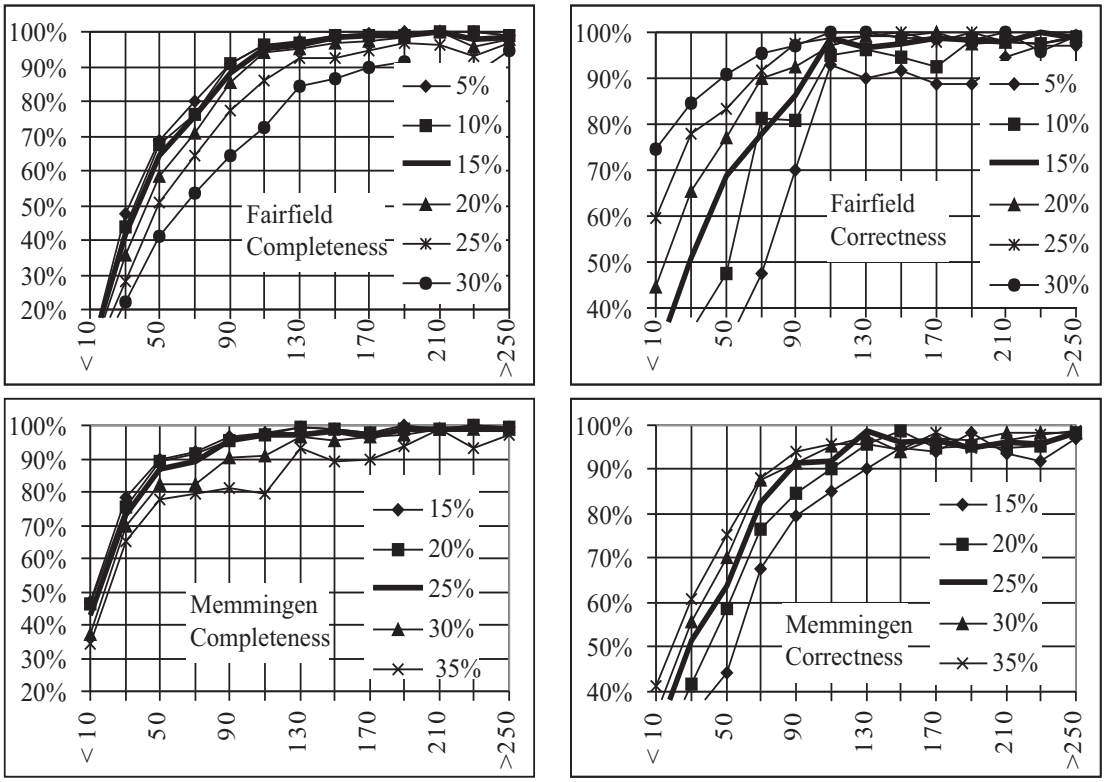


Figure 5.11: Completeness and correctness of the results for different a priori estimates of the percentage of tree pixels P_T [%] as a function of the building size [m^2].

5.5.4 Influence of the Sensor Resolution

Figure 5.12 shows the influence of the resolution of the sensors involved on the classification results. The variants were computed using different values for the grid width Δ of the DSMs. It clearly shows that the results commence to deteriorate for Δ = 2 m. If Δ = 3 m, they are unacceptable, because more than 30% of the buildings covering an area of about 230 m^2 (25 pixels) or smaller are not detectable. At a sampling distance of 3 m, buildings are no longer characterised by low values of surface roughness, because the areas used to compute surface roughness become so large that they include both roof and terrain points. We conclude from the results presented in figure 5.12 that in order to detect buildings of a size of 100 - 150 m^2 reliably from ALS data and multi-spectral images, the resolution of these data must be at least 1.5 m.

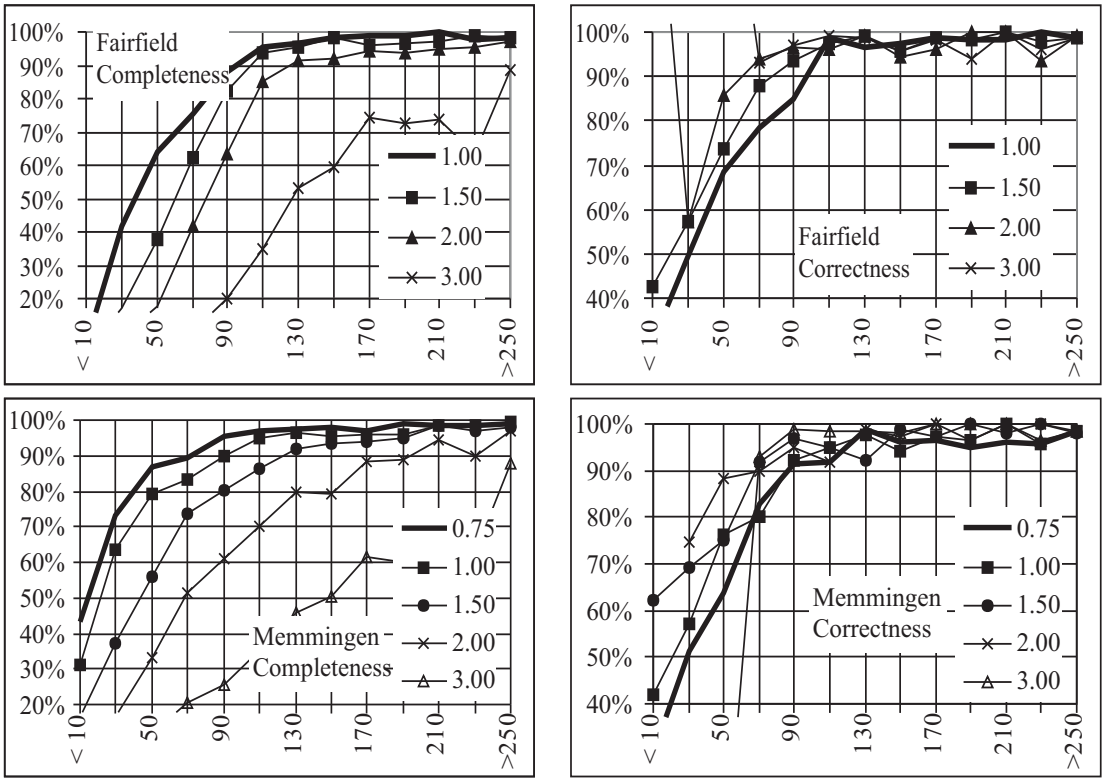


Figure 5.12: Completeness and correctness of the results for different grid sizes Δ [m] of the DSM as a function of the building size [m^2].

5.5.5 Contributions of the Individual Classification Cues

In order to evaluate the contributions of the individual sensors to the accuracy of building detection, we apply our method for building detection to four different combinations of sensors. In the variant LP only the three cues ΔH, R, and D are used, and they are derived from last pulse ALS data. In variant $LP + FP$, we combine first pulse data with last pulse data and thus use the cues ΔH, R, D, and ΔH_{FL}. In variant $LP+NDVI$, we use last pulse data and the NDVI, thus ΔH, R, D, and NDVI. The last variant, All, uses all available data. For Memmingen, different values for the percentage P_T of the scene covered by trees had to be used for the variants, depending on whether the first pulse or the last pulse data were used for surface roughness. Thus, for variants $LP + FP$ and All, the standard value $P_T = 25\%$ was used, whereas for variants LP and $LP + NDVI$, we used $P_T = 15\%$.

Figure 5.13 shows both completeness and correctness achieved for these four variants in both data sets. In Fairfield, completeness is almost identical for all variants. The variants without the NDVI perform better by about 10% for buildings with an area between 50 m^2 and 70 m^2. In Memmingen, variant LP performs worse than the others for buildings smaller than 190 m^2, with up to 30% of the buildings missing. The other variants differ by up to 10% for buildings covering an area of 90 m^2. A certain improvement is achieved by adding the first pulse data to the last pulse data; adding the NDVI increases completeness further for buildings larger than 50 m^2, whereas for smaller buildings, variant $LP + NDVI$ performs best. The correctness shows larger differences between the variants. In Fairfield, variants LP and $LP + FP$ perform almost equally as well. Adding the NDVI to LP increases correctness by about 1-3% for buildings larger than 130 m^2, but by 10%

to 20% for buildings smaller than that limit. In Memmingen, variant $LP + FP$ performs worse than the others, while variant LP is better, but this is contrasted by the unacceptable completeness values. Variant $LP + NDVI$ and the variant using all data perform almost equally well. Adding the NDVI increases the correctness by 5% - 15%, with the increase growing with decreasing building size. To sum up, in all variants for both data sets with the exception of variant LP for Memmingen (which however has unacceptable completeness values), the results are almost the same for buildings larger than 150 m^2. However, as the limits of classification by ALS data are reached for smaller buildings, the NDVI increases the correctness of the results by up to 20%.

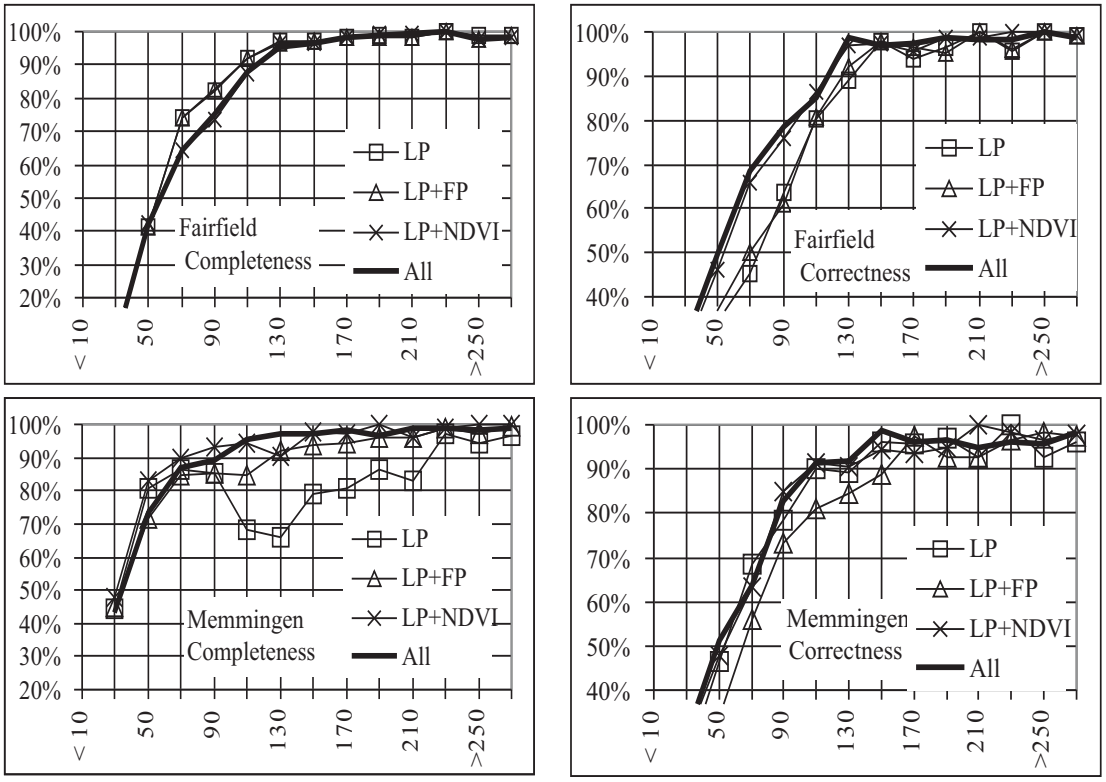


Figure 5.13: Completeness and correctness of the detection results of the building size [m^2] for the four variants described in the text [%]. Above: Fairfield, below: Memmingen.

5.6 Conclusion

We have presented a method for building detection based on Dempster-Shafer fusion of ALS data and multi-spectral images. We have validated the assumptions of the model for assigning probability masses using two data sets comprising both different sensor and scene characteristics. For the pixel-based classification we found simple rules for setting the parameters of that model if an estimate for the area covered by trees is known. This was made possible by a re-parameterisation of the model for surface roughness. We have further improved the method by considering the uncertainty of the NDVI and by post-classification. We have also evaluated the method, giving detailed quality measures on a per-pixel and on a per-building basis. Our experiments show clearly how the results of building detection depend on the building size. Buildings larger than 120 m^2 could be reliably

detected in both data sets. Most buildings between 50 m^2 and 120 m^2 are detected, whereas buildings smaller than 30 m^2 were not able to be detected.

We assessed the sensitivity of the results to the parameter settings in the classification process. The parameters for the probability masses have to be known relatively precisely if excellent results are to be achieved for buildings smaller than 150 m^2; otherwise, good results can still be achieved, but the full potential of the method for detecting small buildings will not be exploited. The parameters for the NDVI have to be determined in a training phase. For the height differences between the DSM and the DTM, and for the height differences between the first and the last laser pulses, the standard parameter settings are well-chosen and will usually not need adaptation. The most difficult parameter to choose is the percentage P_T of trees which is used to derive the probability masses of the surface roughness.

Our experiments also show the limitations of building detection based on the sensor resolution. In order to achieve good results for residential buildings covering an area of about 100 m^2 to 150 m^2, the sensor resolution must be at least 1.5 m. At a resolution of 3 m, only large structures can be detected reliably. An investigation into the contributions of the individual cues showed that the main contribution of the NDVI is to increase the correctness by up to 20% for small to medium-sized buildings. First pulse data also help, though to a lesser degree.

Acknowledgements

This work was supported by the Australian Research Council (ARC) under Discovery Project DP0344678 and Linkage Project LP0230563. The Fairfield data set was provided by AAMHatch (www.aamhatch.com.au). The Memmingen data set was provided by TopoSys (www.toposys.com). The authors want to thank the students Christian Eberhöfer, Werner Mücke, and Gerhard Summer for carrying out the experiments described in this chapter.

Chapter 6

Automatic Generation of High-Quality Building Models From LIDAR Data[1]

6.1 Introduction

Automating data acquisition for 3D city models is an important research topic in photogrammetry. In addition to techniques that rely on aerial images, generating 3D building models from point clouds provided by *Li*ght *D*etection *A*nd *R*anging (LIDAR) sensors is gaining importance. The progress in sensor technology has triggered this development. Today, airborne laser scanners can deliver dense point clouds with densities of up to one point per square meter. Using this information, it is possible to detect buildings and their approximate outlines and also to extract planar roof faces and create models that correctly resemble the roof structures.

This chapter presents a new method for automatically creating polyhedral building models without using ground plans. The ability to group neighbouring planes is still a work in progress. My development team (referred to in this chapter as "we" [2]) computed the examples presented in this chapter using the LIDAR data from a test site in Vienna captured by the TopoSys company. The resolution of the original point cloud is 0.1 m (in-flight) by 0.9 m (cross-flight). From that point cloud, we derived a regular grid for building extraction. The test data were captured in the course of a pilot project for the municipality of Vienna to evaluate and compare various techniques for generating 3D city models. My initial results show the high potential of this method.

The most recent achievements in the field of automated acquisition of 3D building models are based on integrating data from two or more sources - such as LIDAR and digital aerial images [Ameri, 2000a] - to overcome the drawbacks of specific sensor types. The trend toward combining data from multiple sensors for automatically reconstructing topographic objects is triggered by the fact that new sensor types have become available. In the future, the task of data acquisition for geographic information systems could be performed by "multisensor-grammetry" rather than by traditional photogrammetry. We hope this chapter contributes to the development of this "multisensor-

[1] © IEEE. Reprinted, with permission, from Rottensteiner, F., Automatic generation of high-quality building models from LIDAR data, IEEE Computer Graphics and Applications 23(6), pp. 42-51, 2003.

[2] This was inserted by the journal editors, who did not like the author's plural used throughout this chapter. The stylistic peculiarities of this chapter compared to the remainder of this work are also the result of this editing process, since an "active" writing style was required.

grammetry" by describing how aerial images could be integrated into the workflow of building extraction.

6.2 Recent Work on Building Extraction

There have been several attempts to detect buildings using LIDAR data. One approach is to classify the LIDAR points according to whether they belong to the terrain, to buildings, or to other object classes. Researchers commonly use morphological opening filters or rank filters to determine a digital terrain model that is subtracted from the digital surface model. By applying height thresholds to the normalized surface model, researchers have obtained an initial building mask [Weidner, 1997]. The algorithm we use for building detection from LIDAR points is based on the method for terrain model generation by robust interpolation [Rottensteiner and Briese, 2002].

Research teams have tackled the geometrical reconstruction of buildings in previously detected regions of interest in two ways. First, you can instantiate parametric primitives and fit them to the data if you find sufficient evidence. Second, you can detect planar segments in a surface model created from the LIDAR points and derive polyhedral building models by grouping these planar segments. Because parametric primitives often have a rectangular footprint, you can use them if 2D ground plans offer a precise location of the building outlines. In these cases, you split the polygon delineating a building in a 2D map into rectangular regions. In each rectangle, you can determine the parameters of the models using the surface model, and then accept the model that achieves the best fit [Brenner, 2000, Vosselman and Dijkman, 2001].

The data-driven generation of polyhedral building models from LIDAR data only makes sense if the point density is high enough so that you can locate a sufficient number of data points at least in the most relevant planes of the roofs. Ground plans can reduce search space for estimating the parameters of adjoining planar segments because the gradient direction of such planes is usually perpendicular to the adjacent polygon segment in the ground plan [Haala et al., 1998]. You can find initial planar segments by segmenting the surface model.

One research team described four different range image segmentation algorithms that were based on such techniques as region growing, clustering, and an analysis of scan lines [Hoover et al., 1996]. This team developed a framework for evaluating segmentation algorithms and applied it to the four segmentation algorithms in an extensive test with data from structured light and laser range-finder images. The team concluded that segmentation still can be improved because their algorithms have problems with small regions and with correctly detecting the borders of adjacent regions.

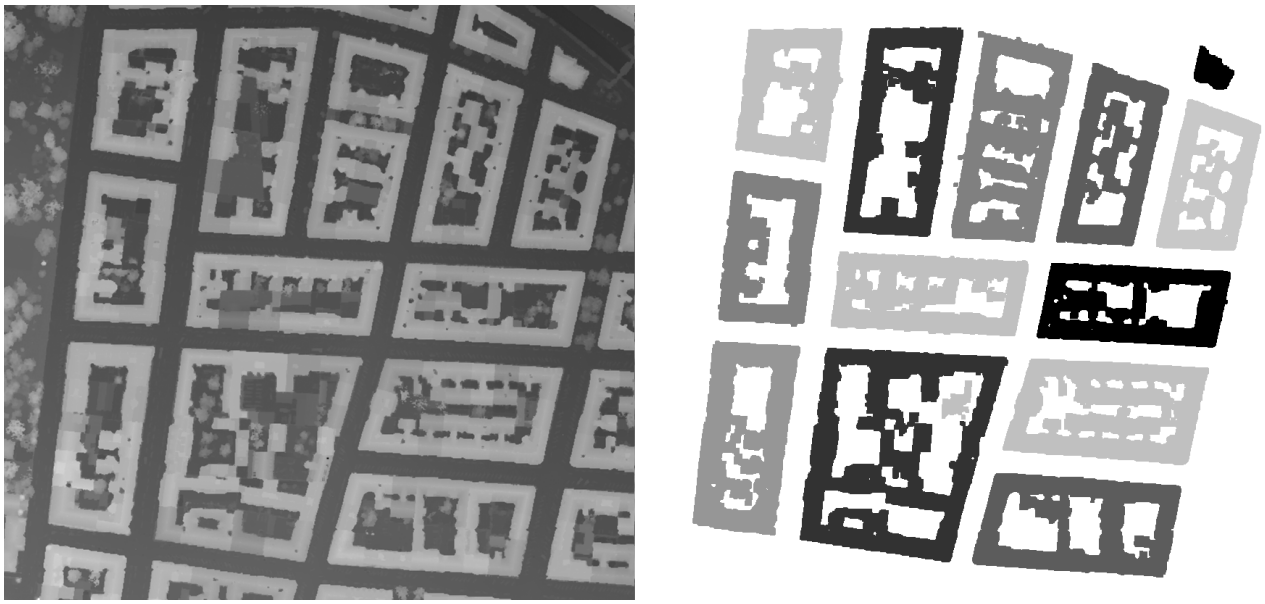
Another research effort applied a region-growing algorithm to LIDAR data using a similar framework for evaluation [Geibel and Stilla, 2000]. In yet another technique, the team grouped the neighbouring segments as soon as they found initial planar segments, which involved finding consistent intersections at the building vertices [Baillard et al., 1999]. In addition, they combined the 3D border polygons to obtain consistent building models. The approach requires adding vertical walls and the floor to the model. They applied a coarse-to-fine strategy by first searching for the most relevant structures in the data and using refined methods for modelling the buildings in regions not being explained sufficiently by the initial models. After generating the building model, they project its roof edges back to aerial images where the model edges are matched with image edges. This technique increases the accuracy of the model, especially with respect to the building outlines. In the

past, we have used such a model-driven technique for wire frame fitting for measuring parametric primitives in semiautomatic building extraction [Rottensteiner, 2001]. Because it is general enough to be applicable to any polyhedral model, we use it in the context of this work.

6.3 Method Overview

Building extraction consists of two steps: detection and reconstruction. You have to detect building candidate regions in the digital surface model (DSM) by a classification of the LIDAR points. As a result, you obtain regions of interest for the geometric reconstruction of the buildings. In the subsequent processes, our system handles these regions of interest individually. Furthermore, in the regions of interest thus detected, the buildings have to be reconstructed geometrically, which results in 3D polyhedral models.

The system for building detection applies a two-step procedure for classifying LIDAR points. First, it uses linear prediction to separate terrain points from off-terrain points hierarchically, starting from thinned-out data. It derives two digital elevation models of identical grid width, and it computes a digital terrain model (DTM) from the points classified as terrain points with a high degree of smoothing. The system also computes a DSM from all points without smoothing, as shown in figure 6.1(a). Second, the off-terrain points have to be classified to separate points on buildings from points on other objects. The system performs this by thresholding the height differences to find initial building regions. The system improves these initial results by morphologic filtering and by a texture analysis to remove vegetation areas, as shown in figure 6.1(b). We have described this method in more detail in [Rottensteiner and Briese, 2002].



(a) (b)

Figure 6.1: Building detection in a test site in Vienna: (a) digital surface model and (b) building regions.

Having detected the building regions, the workflow for the geometric reconstruction of the

buildings consists of four steps.

1. *Detection of roof planes.* In building candidate regions, the system applies a segmentation on the basis of an analysis of the variations of the DSM normal vectors to find planar segments that are expanded by region growing algorithms.

2. *Grouping of roof planes and model generation.* Neighbouring roof segments, being coplanar, have to be merged. As soon as the system finds no more hypotheses for grouping planar segments, the system completes the resulting polyhedral model.

3. *Consistent estimation of the model parameters.* Having generated initial building models, their geometric parameters must be improved by a consistent estimation procedure taking into account all the available sensor information.

4. *Model regularization.* Finally, the system improves the models by introducing hypotheses about geometric constraints between planes, and it repeats parameter estimation.

The resolution of the LIDAR data - which is still below the resolution of aerial images - limits the level of detail of the building models and the accuracy of the positions of step edges.

6.4 Detection of Roof Planes

One research team describes a method for polymorphic feature extraction. This method aims at classifying texture as being homogeneous, linear, or point-like by an analysis of the first derivatives of a digital image [Fuchs, 1998]. That classification is based on a significance test of the squared gradient norm of the digital images, the most important parameter being the significance level α of that test, from which the threshold for classification is derived.

The system applies this method to the first derivatives of the DSM. Pixels classified as homogeneous are surrounded by pixels having similar components of the surface normal vectors. That is, they are in a region containing coplanar points. Figures 6.2(a) and 6.2(b) show the binary image of homogeneous pixels in one of the building regions from figure 6.1, derived by using two different significance levels. The system applies a connected component analysis to the binary image of homogeneous pixels to detect seed regions for region growing. However, because of classification errors, either too few such regions are detected (figure 6.2(a)) or the detected regions turned out to be too large (figure 6.2(b)).

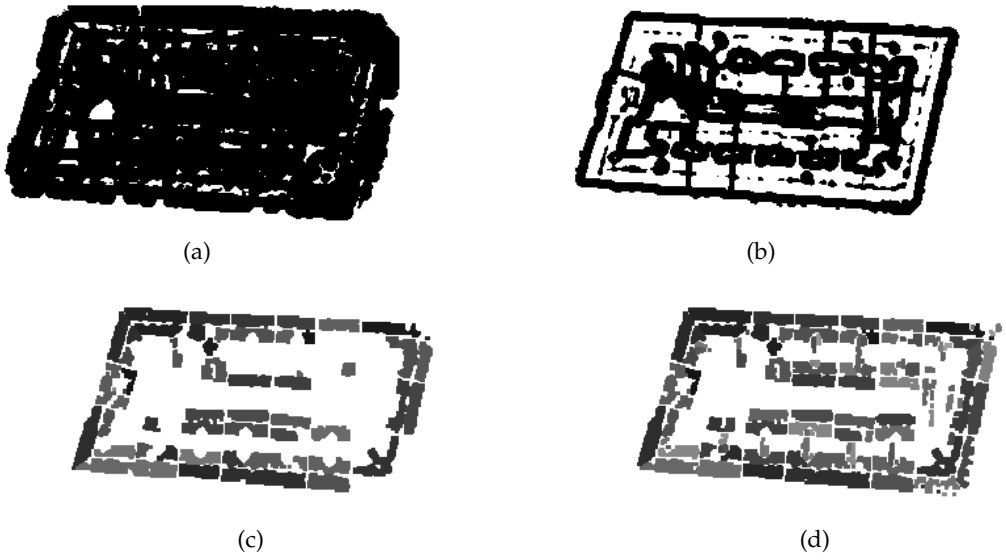
To avoid these segmentation errors, we propose an iterative strategy for roof-plane detection, splitting the connected components of homogeneous pixels into smaller parts by morphologic filtering and only allowing well-fitting planes to grow. Although the results thus achieved were already quite satisfying, the procedure failed if the initial segments were shaped in a way that parts belonging to different planes in object space could not be separated.

We improved the method to give it a better statistical basis. Again, the system performs roof-plane detection iteratively, but it iterates over different significance levels in texture classification. The system starts by texture classification using a tight threshold, and thus a high significance level. Statistically, this means that the system accepts a large percentage of homogeneous pixels erroneously

classified as non-homogeneous, and a small percentage of non-homogeneous pixels erroneously classified as homogeneous. Only the pixels in the most significantly planar regions are actually classified as homogeneous (figure 6.2(a)).

The system uses connected regions of homogeneous pixels as seed regions for region growing, allowing only regions achieving a good planar fit to grow. New pixels adjacent to a region are added if their distances from the adjusting plane is below a certain threshold. The system repeats this step until no additional pixel can be added to any segment. Figure 6.2(c) shows the segment label image derived from growing the seed regions from figure 6.2(a). The system has already assigned about 50% of the pixels of the building region to one of these planes, each plane having an error of planar fit better than $\pm$15 cm. After that, the system performs the classification again using another threshold and applying our procedure of seed region selection to connected components of pixels now classified as homogeneous, but not yet assigned to a planar segment in the previous iteration.

In this way, the system performs a certain number of iterations, each iteration finds new planar segments and accepts less and less significantly homogeneous areas to take part. From a practical point of view, you can select the threshold of the first iteration so that the system uses only the most significantly planar homogeneous pixels. Figure 6.2(b) shows the classification results of the eighth iteration, and figure 6.2(d) shows the resulting segment label image. All the detected segments have an error of planar fit better than $\pm$15 cm. About 64% of the pixels in the building region are assigned to one of these roof planes.



(a) (b)

(c) (d)

Figure 6.2: Detection of planar segments from a digital surface model in one of the building regions from figure 6.1. (a) and (b) show classification results of polymorphic feature extraction using different significance levels (first and last iteration). (c) Planar regions obtained by applying a connected component algorithm and region growing to (a). (d) Segment label image after the last iteration. A colour version of this figure can be found on page 130.

Figure 6.3(a) shows the unclassified pixels of the building region corresponding to figure 6.2(d). Mostly, the unclassified pixels are those at the borders of the planar regions, especially at the positions of step edges or chimneys. However, there are also larger patches of unclassified pixels corresponding either to small roof structures not yet detected, to roof parts not planar, and to regions that are

(a) (b) (c)

Figure 6.3: Postprocessing of the segmentation results: (a) unclassified pixels in the building region; (b) segment label image after post-processing containing additional planar segments; and (c) unclassified pixels after postprocessing. A colour version of this figure can be found on page 130.

not parts of the building.

We can improve this segmentation by post-processing, trying to find additional planes in the unclassified regions, and eliminating trees and other objects not belonging to the building. First, the system searches for connected components of unclassified pixels having a small height gradient strength. This is motivated by the fact that the kernel size for computing the height gradient is smaller than the size required for the texture classification described previously, so that it is possible to find smaller seed regions. The system applies region growing to these regions, allowing larger residuals, so that the additional planar segments have a worse planar fit than the original ones.

Thus the system might get segments generalizing the actual shape of a building part that cannot be reconstructed with more details, given the resolution of the LIDAR data. Figure 6.3(b) shows the final segment label image containing the additional planes. Finally, the system searches for connected components of unclassified pixels having a great percentage of pixels classified as point-like in polymorphic feature extraction, which is an indicator for vegetation areas. Figure 6.3(c) shows the unclassified pixels after having eliminated these areas from the building region.

At this point, only DSM pixels at the borders of the planar segments remain unclassified. Table 6.1 gives a summary of the results of roof plane detection for all the building regions detected in the test site shown in figure 6.1. Altogether 69.9% of all pixels classified as building pixels are assigned to a planar segment; 92.1% of these pixels (or 64.4% of all building pixels) are in a planar segment having an error of planar fit better than ± 15 cm. The remaining 7.9% of the pixels assigned to a planar segment mostly correspond to planes added to the model in the second segmentation phase, generalizing more detailed building shapes.

6.5 Grouping of Roof Planes and Model Generation

First, we derive the neighbourhood relations of the planar segments from a Voronoi diagram that is based on a distance transformation of the segment label image: Each pixel inside the region of interest not yet assigned to a planar segment is assigned to the nearest segment (figure 6.4(a)). After finding the neighbourhood relations, coplanar segments are merged. The coplanarity test is based on a Fisher test comparing the root mean square (RMS) errors of planar fit obtained by applying two different mathematical models: two separate planes versus one compound plane. Figure 6.4(b) shows the Voronoi diagram after merging coplanar segments.

RMS [m]	Planes [%]	All [%]
0.00-0.05	19.5	13.6
0.05-0.10	57.3	40.1
0.10-0.15	15.3	10.7
0.15-0.20	6.5	4.5
0.20-0.25	0.8	0.7
0.25.0.30	0.3	0.2
0.30-0.35	0.2	0.1
0.35-0.40	0.0	0.0
0.40-0.45	0.1	0.0

Table 6.1: Distribution of the root mean square errors of the planar fit.

(a) (b)

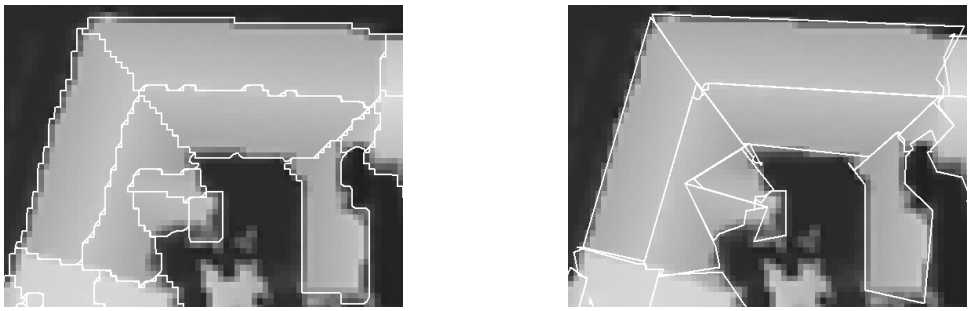
Figure 6.4: (a) Voronoi diagram of the label image in figure 6.3(b) and (b) Voronoi diagram after merging coplanar segments. A colour version of this figure can be found on page 130.

We then analyze the neighbourhood relations of the remaining segments. The boundary polygons of the Voronoi regions give a first estimate for the planimetric positions of the segment outlines (figure 6.5(a)). We need to classify each portion of the boundary separating two planar segments according to the geometric configuration of these segments. There might be an intersection, a step edge, or both an intersection and a step edge.

The system computes all the intersection lines. For a pair of neighbouring planar segments, the system checks whether the error of the original border polygon is below a certain threshold with respect to the intersection line. If this is the case, the intersection line acts as the actual border between these planar segments, and the system updates the border polygons of the two planar segments to contain that intersection line. If this is not the case, the system assumes a step edge. It tries to locate its planimetric position precisely by searching for the maximum of the gradient strength of the LIDAR DSM in the direction of the normal vector of the original border polygon in that area. In the third case, not yet considered by the current implementation of this method, the border polygon between two planar segments has to be split into smaller parts.

Having improved the shapes of the border polygons of the roof segments, the polyhedral models have to be created. The planar segments become the roof faces of the model, the vertices of the boundary polygons become the building vertices, and the edges of these polygons become the edges of the polyhedral model. A vertical face corresponding to a wall is inserted for each polygon edge classified as a step edge.

The modules for grouping and model generation are still works in progress. Figure 6.5(b) shows preliminary results of computing intersections and step edges. We must improve the search



(a) (b)

Figure 6.5: Part of the roof polygons of the building in figures 6.2, 6.3, and 6.4 superimposed on the digital surface model: (a) the original pixel chains from figure 6.4(b) and (b) after finding intersections and step edges.

strategy for step edges, implement the third case of mutual relations between planar segments, and improve the boundary polygons at the building vertices, because there are still short polygon edges not part of an intersection line or getting weak support from step edge detection.

Figure 6.6 shows the reconstructed roof polygons superimposed to an aerial image of scale 1:7000. We could not reconstruct smaller roof details, such as the chimneys, given the resolution of the LIDAR data. There are still planimetric offsets in the images that correspond to about five image pixels. We did not consider short polygon segments for step edge detection, which leads to their somewhat ragged appearance. Figure 6.7 shows a visualization of a 3D model created from vertical prisms bounded by the roof polygons and the floor. These preliminary results are encouraging because they show the potential of high-resolution LIDAR data for building extraction.

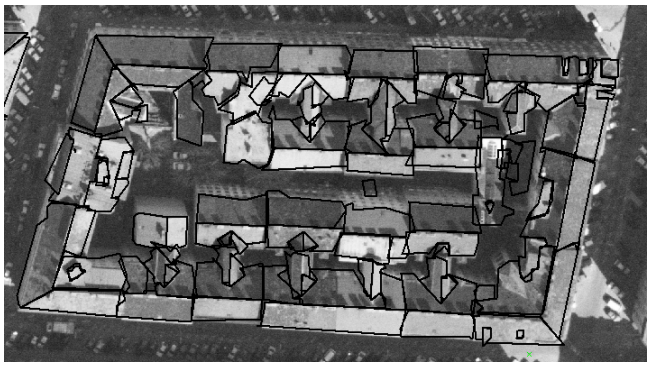


Figure 6.6: Roof polygons of the building in figures 6.2, 6.3, and 6.4 back projected to an aerial image. Ground sampling distance: 17 cm.

6.6 Consistent Estimation of the Model Parameters

In the previous phases of building reconstruction, we determined the parameters of the planar segments individually. We determined the building vertices from the intersections of adjacent planes or

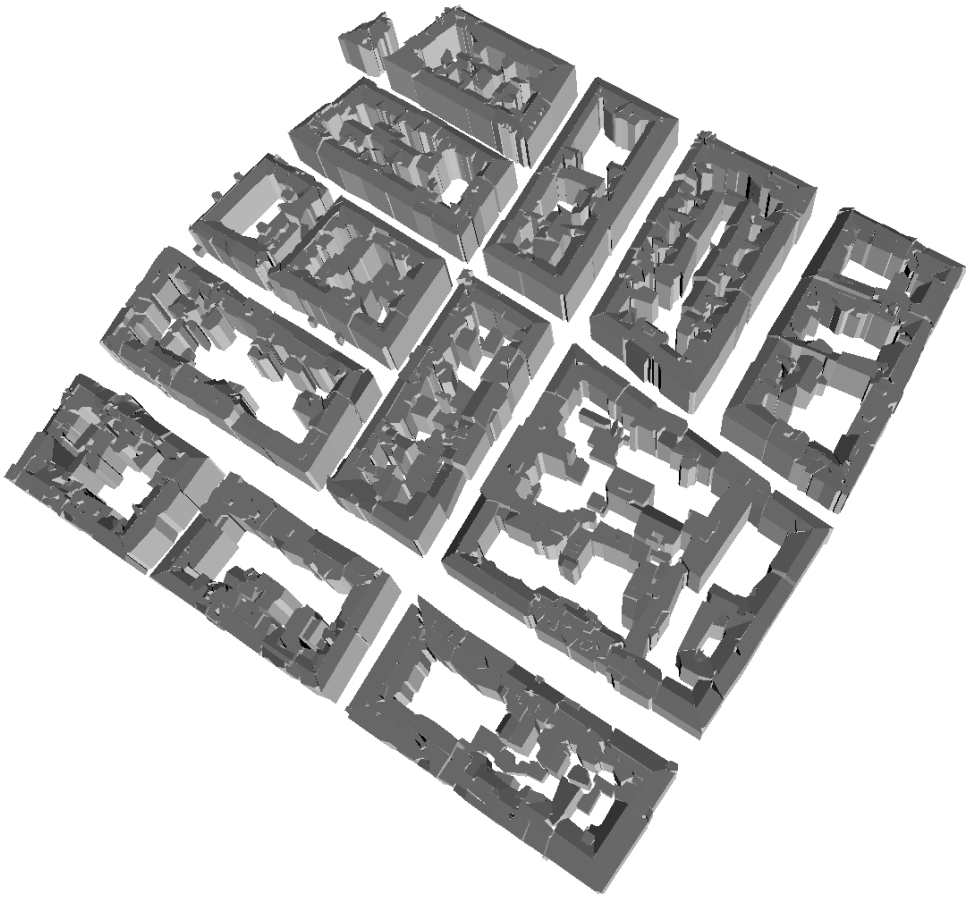


Figure 6.7: VRML visualization of a model created from vertical prisms bordered by the boundary polygons and the floors.

by step-edge extraction. An overall adjustment including all available sensor information is important to get geometrically consistent building models and to consider geometric constraints.

We use our hybrid adjustment program *ORIENT*, for estimating the model parameters [Kager, 1989]. Our method finds a mapping between the boundary representation (B-rep) of a building and a system of shape observations (referred to as *GESTALT* observations) representing B-rep and the geometric constraints imposed by that modeling technique. In this model, a *GESTALT* observation is the observation of a point $\mathbf{P}$ situated on a polynomial surface. The polynomial is parameterised in an observation coordinate system (u, v, w) related to the object coordinate system by a shift $\mathbf{P_0}$ and three rotations $\mathbf{\Theta} = (\omega, \varphi, \kappa)^T$. The actual observation is a fictitious observation of the largest component of the distance vector to be 0. Using the shorthand $(u_R, v_R, w_R)^T = \mathbf{R}^T(\mathbf{\Theta}) \cdot (\mathbf{P} - \mathbf{P_0})$, with $\mathbf{R}^T(\mathbf{\Theta})$ a transposed rotational matrix parameterized by $\mathbf{\Theta}$, and restricting ourselves to vertical planes for walls and generally tilted planes for roofs, there are three possible formulations of *GESTALT* observation equations:

$$0 + r_u = \quad m_u \cdot u_R + a_{00} + a_{01} \cdot m_v \cdot v_R \tag{6.1}$$

$$0 + r_v = \quad m_v \cdot v_R + b_{00} + b_{10} \cdot m_u \cdot u_R \tag{6.2}$$

$$0 + r_w = \quad m_w \cdot w_R + c_{00} + c_{10} \cdot m_u \cdot u_R + c_{01} \cdot m_v \cdot v_R \tag{6.3}$$

In equations 6.1, 6.2, and 6.3, r_i are the corrections of the fictitious observations of coordinate i and $m_i \in \{-1, 1\}$ are mirror coefficients. A *GESTALT* is a set of *GESTALT* observations for all vertices neighbouring one specific face of the building's B-rep. The *GESTALT* is the representation of that face in adjustment, its surface parameters (either a_{jk}, b_{ik}, or c_{ij}) being the parameters of the planar face, and the *GESTALT* observations connect these parameters with the coordinates of the vertices $\mathbf{P}$.

For each face of the B-rep of the building model, such a *GESTALT* is defined and parameterized in one of the three ways shown in equations 6.1-6.3, either equation 6.1 or 6.2 being used for walls and equation 6.3 for roof faces. Note, however, that an application is free to decide which of the parameters in equations 6.1-6.3 ($\mathbf{P}$, $\mathbf{P_0}$, $\mathbf{\Theta}$, surface parameters a_{jk}, b_{ik}, or c_{ij}) are to be determined in adjustment and how to parameterise a surface. In addition, different *GESTALT*s can refer to identical transformation or surface parameters. You can use these properties - and a proper selection of the mirror coefficients - to enforce geometric constraints.

The estimation of building parameters from sensor data by representing the B-rep of buildings in adjustment by a system of *GESTALT*s has originally been applied to parametric primitives in semiautomatic building extraction [Rottensteiner, 2001]. Figure 6.8 shows a building primitive resembling a saddleback roof. Table 6.2 provides an overview of the parameterization of the seven *GESTALT*s corresponding to the seven faces of the primitive. The primitive is modeled to be symmetric with respect to the vw-plane and to have a rectangular footprint. All *GESTALT*s refer to the same observation coordinate system centred in the center of the floor and having a vertical w-axis, thus $\omega = \varphi = 0$. Although the building is modelled in B-rep, only four pose ($\mathbf{P_0}$ and κ) and four shape parameters (b^f_{00}, a^r_{00}, c^r_{00}, c^r_{10}) have to be determined from sensor data. By properly selecting the parameterisations of the *GESTALT*s, we succeeded in reducing the number of parameters to a minimum, thus imposing geometric constraints to the primitive.

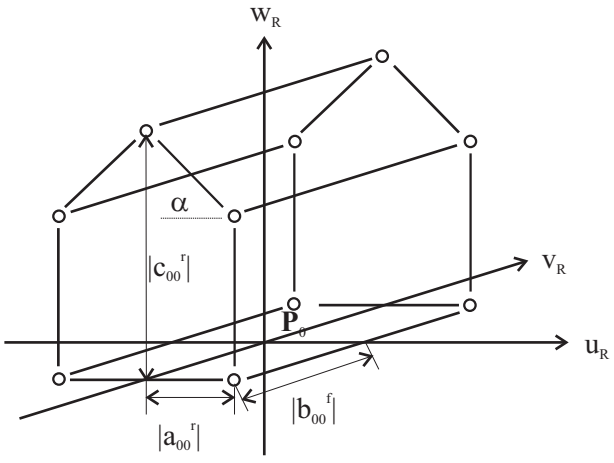


Figure 6.8: Saddleback roof and its parameters.

For building extraction from LIDAR, we don't want to use parametric primitives because doing so could reduce the applicability of the algorithm to rectangular buildings [Vosselman and Dijkman, 2001]. Thus, we use the more general equations for the parameterization of the *GESTALT*s, introducing geometric constraints only where evidence for their occurrence is found. As in equation 6.1, there are dependencies between the constant parameters and $\mathbf{P_0}$ and between

Face Number	Face	Face Equation	Symmetric Face Number	m_i
1	Floor	$0 + r_w = w_R$	-	-
2	Front wall	$0 + r_v = v_R + b^f_{00}$	-	-
3	Rear wall	$0 + r_v = -v_R + b^f_{00}$	2	m_v
4	Right wall	$0 + r_u = u_R + a^r_{00}$	-	-
5	Left wall	$0 + r_u = -u_R + a^r_{00}$	4	m_u
6	Right roof	$0 + r_w = w_R + c^r_{00} + c^r_{10} \cdot u_R$	-	-
7	Left roof	$0 + r_w = w_R + c^r_{00} - c^r_{10} \cdot u_R$	6	m_u

Table 6.2: Overview of the parameterization of the seven *GESTALT*s. m_i is the mirror coefficient that equals -1.

the linear terms and the rotational angles, we declare both $\mathbf{P_0}$ and $\mathbf{\Theta}$ to be constant. For numerical reasons, we situate $\mathbf{P_0}$ inside the building. The rotations are set to 0, which means that the tilts of the planes are modeled by the linear terms of equations 6.1-6.3.

As described previously, for each face of the B-rep of the building model, we define a set of *GESTALT* observations, taking equations 6.1 and 6.2 for walls and equation 6.3 for roofs. The unknowns are the object coordinates of each point $\mathbf{P}$ and the plane parameters (a_{jk}, b_{ik}, c_{ij}). As each building vertex is neighboured by at least three faces, the object coordinates of the vertices can be determined from *GESTALT* observations for these vertices. We use the sensor data to determine the unknown plane parameters for all the planes of the building. We assign LIDAR points to roof planes. For each LIDAR point, we introduce one *GESTALT* observation as well as three direct observations for the point's LIDAR coordinates to avoid singularities with respect to the object coordinates $\mathbf{P}$.

We introduce direct observations for the planimetric coordinates of building vertices at step edges to make the parameters of the wall faces determinable. These direct observations are the results of step edge detection. They are situated at the positions of the maximum gradient strength of the DSM.

Using this mathematic model, we get a highly redundant adjustment system. It is possible to apply robust estimation to eliminate false or contradicting observations. The weights of the observations depend on their a priori standard deviations. The stochastic model's most important parameter is the a priori standard deviation of the *GESTALT* observations. Typically, we select it in the range of a few centimetres.

6.7 Model Regularization

Up to now, we have not yet used any information concerning geometric regularities to improve the quality of the building model, which leads to irregular shapes. In the future, we will develop a constraint generator that has to analyze the polyhedral model and introduce geometric constraints where sufficient evidence for their occurrence is found. If two neighbouring planes ε_1 and ε_2 are found to fulfil a geometric condition, we will add *GESTALT* observations, taking advantage of specific definitions of the observation coordinate system and specific parameterisations of the planes. Figure 6.9 shows three such constraints.

In all cases, one of the axes of the observation coordinate system is the intersection of ε_1 and ε_2 and one of the vertices of the intersection line is the reference point $\mathbf{P_0}$ of both planes. The rotations

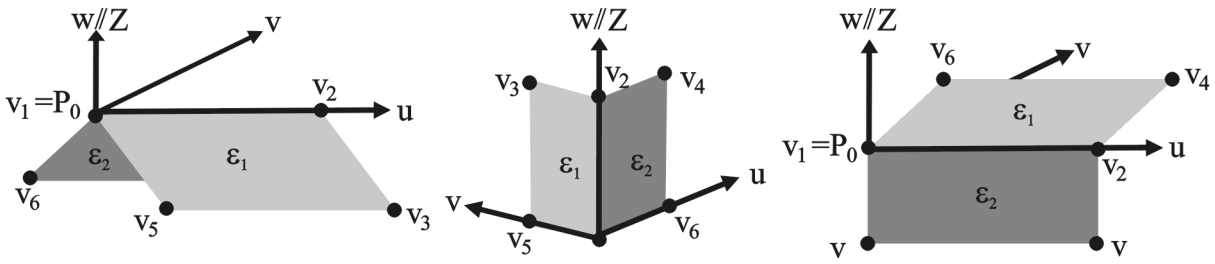


Figure 6.9: Three possible geometrical constraints between two planes ε_1 and ε_2. Left: a horizontal ridge; centre: two orthogonal walls; right: a horizontal eave.

ω and φ are 0 and constant, but there is one additional unknown rotational angle κ. For each vertex $\mathbf{v_i}$ of the planes, we add *GESTALT* observations for ε_1 and/or ε_2. We parameterize these planes in specific ways:

- The intersection of two planes is a horizontal ridge: $\varepsilon_1 : 0 + r_w = w_R + c_{01}^1 \cdot v_R$ and $\varepsilon_2 : 0 + r_w = w_R - c_{01}^1 \cdot v_R$. There is only one tilt parameter c_{01}^1. Symmetry is enforced by selecting $m_v = -1$ for ε_2.
- Two walls are perpendicular: $\varepsilon_1 : 0 + r_u = u_R$, $\varepsilon_2 : 0 + r_v = v_R$. There is no additional surface parameter to be determined.
- A wall and a roof plane intersect in a horizontal eave: $\varepsilon_1 : 0 + r_w = w_R + c_{01}^1 \cdot v_R$ and $\varepsilon_2 : 0 + r_v = v_R$. There is an additional unknown roof tilt c_{01}^1.

The a priori standard deviations of the *GESTALT* observations describe the stochastic model of these constraints. After creating these additional observations, the estimation of the model parameters has to be repeated.

6.8 Integration of LIDAR and Aerial Images

The quality of the building models derived from LIDAR data is restricted by the LIDAR sensor's ground resolution. As the geometric resolution of aerial images is still much better than that, it would be best to integrate aerial images into the workflow of building extraction. As we have seen in the previous sections, the most relevant roof structures can be extracted well from LIDAR data. There are two stages where the aerial images can help improve the quality of the resulting models. They can help detect additional planar segments in model generation and they can be used to improve the geometric quality of the model edges by matching these model edges with grey level edges extracted in the digital images.

6.8.1 Detection of Planar Segments

There are still some unclassified pixels, and in the post-processing phase we have accepted segments having RMS errors planar fit that indicate that small roof structures have been generalized. In both

cases, it's possible to improve the segmentation results using the information provided by the aerial images.

We use polymorphic feature extraction for segmenting the digital images. This technique results in a set of segments of homogeneous grey levels for each image. Figure 6.10 shows one of the areas containing unclassified pixels in the LIDAR DSM, segmentation results in a corresponding aerial image, and a sketch of the actual roof structure. In the LIDAR DSM, that area has a width of about five pixels. It is impossible to separate the two roof planes obviously connecting two larger building parts on the upper and the lower margins of the figure because the sizes of the filters involved in seed region selection limit the size of detectable planes. However, in both aerial images, the grey level segmentation gives two segments corresponding to two roof planes. Thus, figure 6.10 shows that it makes sense to use the image segmentation results in the areas where no planes can be detected in the LIDAR data.

(a) (b) (c)

Figure 6.10: (a) Small portion of the DSM in one of the unclassified areas in figure 6.3(a), (b) segmentation results of the same area in an aerial image, and (c) sketch of the roof structure.

It is possible to project the image segments to the DSM to get a first estimate for the plane parameters. In addition, homologous segments from different images can be matched, so that the segmentation results from different images can be combined. In the future, the most significant segments (those obtaining support in multiple images) will be selected for further processing. They can be added directly to the list of planar segments or be used as new seed regions for region growing in areas where no such regions could be derived from the LIDAR data alone due to the restrictions imposed by the sizes of the involved filters.

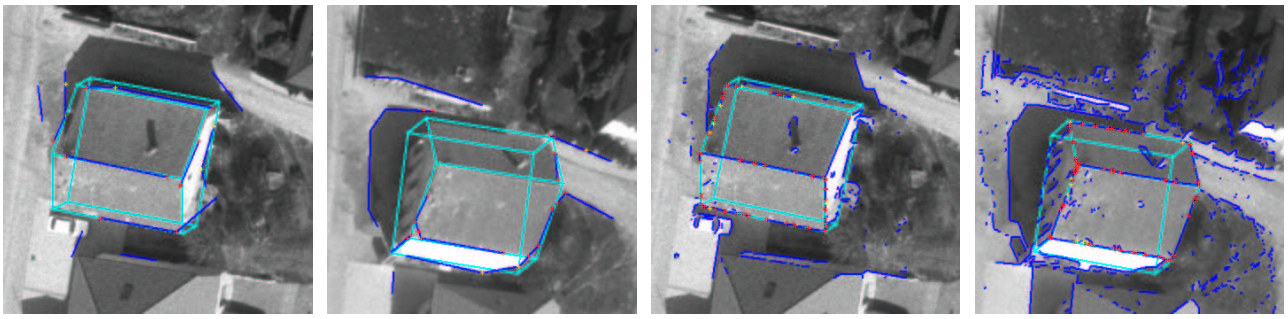
6.8.2 Wire Frame Fitting

Wire frame fitting is a good approach for improving the geometric quality of the polyhedral models created from LIDAR data. Unlike LIDAR data, you can extract the grey level edges corresponding to the building outlines precisely in the images. Projecting the polyhedral model back to the images (figure 6.6), the frame's edges can be matched with image edges. In each image, the projected position of an object edge gives us a region of interest, and we search for image edges inside that region of interest that are almost parallel to the projected object edge.

Image edges fulfilling this condition are supposed to be matching candidates, and they are considered in the estimation of the model parameters by an expansion of the adjustment model. In

addition to the observations described there, the image coordinates of the end points of the image edge segments are assigned to a roof edge. For each end point, we get two image coordinates and two *GESTALT* observations - one for each face that's an object edge neighbour - and three additional unknown object coordinates $\mathbf{P}$.

Whereas we search for matching candidates for all roof edges independently in all images, we perform parameter estimation in an overall robust process in a hybrid adjustment of the *GESTALT* observations representing the object model, the LIDAR points, the positions of the step edges, and the image coordinates of image features. We apply robust estimation to determine false matches between image and object edges. We first used this model-fitting algorithm for automatically measuring parametric building models in a system for semiautomatic building extraction from aerial images. In a test project in the context of semiautomatic building extraction, we have shown that results with an accuracy range of a few centimetres can be achieved using that method [Rottensteiner, 2001] (cf. figure 6.11).



(a) (b) (c) (d)

Figure 6.11: Wire frame fitting of a building primitive resembling a saddleback roof in two images at two resolution levels of the image pyramids. Blue is the extracted image edges, red is the edge segment end points accepted as matches, yellow is the false matches, and cyan is the final positions of the wire frame. A colour version of this figure can be found on page 131.

6.9 Conclusions

This technique can handle polyhedral buildings of arbitrary shape. Unlike other techniques, it doesn't require assumptions about rectangularity of the footprint, nor a priori information about the building outlines. The segmentation algorithm used for roof plane detection belonging to the group based on seed-region selection and region growing. The preliminary results from a test site in Vienna showed the method's high potential: All buildings in the test area could be detected, although their outlines had a rather complex shape. The reconstructed models resembled the roof shapes well according to a visual inspection.

We have also discussed the issue of integrating aerial images into the workflow to obtain better results by what could be called "multisensor-grammetry". In that way, we could try to overcome the problems associated with trying to detect small planar segments in the LIDAR data caused by the limited resolution of the LIDAR sensors. The system described here is still work in progress. Planar segmentation and the mathematic model of hybrid adjustment including *GESTALT* observations do already exist. We have implemented wire frame fitting and tested the technique for parametric

primitives. Whereas the mathematical model for creating geometric constraints has already been elaborated, the module for analyzing the shape of the models for introducing such constraints has not yet been implemented. Future work will include the implementation of the modules still missing and the assessment of quality parameters for the results.

Chapter 7

Automated Delineation of Roof Planes from LIDAR Data

7.1 Introduction

7.1.1 Motivation and Goals

LIDAR data offer a high potential for automated building extraction. Buildings consist of regular surfaces that can be extracted from LIDAR data making use of surface properties such as local co-planarity. One approach to reconstruct buildings from LIDAR data is to segment the data into planes and then to combine these planes to obtain a polyhedral model [Vosselman, 1999, Vögtle and Steinle, 2000, Rottensteiner, 2003, Alharty and Bethel, 2004]. Alternatively, buildings can be reconstructed by parametric primitives, e.g. [Brenner, 2000]. Using parametric primitives reduces the level of detail that can be achieved as the number of primitives is usually small and most have a rectangular footprint. The greatest problem encountered with generic methods for reconstructing polyhedral models is the delineation of the roof plane boundaries. These boundaries correspond to edges in the LIDAR data. In contrast to edges corresponding to the intersection of neighbouring roof planes that can be determined very precisely, step edges are poorly defined. As step edges occur at building outlines, 2D GIS data are often used in combination with LIDAR data to alleviate this problem, e.g. [Brenner, 2000].

If no GIS data are available, the roof boundaries have to be determined from edges extracted from the LIDAR data. The approximate positions of such edges are given by the boundaries of the planar segments that have been extracted from a Digital Surface Model (DSM) created from the LIDAR data. As these positions are not very precise, the determined polygons appear very ragged and regularisation techniques need to be applied. Some algorithms rely on assumptions with respect to the roof shapes, e.g. on all corners being right-angled [Vosselman, 1999], which reduces the level of detail of the resulting models. Another common feature of such algorithms is that they rely on comparing distances to user-defined thresholds for taking decisions regarding geometric constraints. Of course, the setting of such thresholds is a critical issue. Another problem is that the quality of the outlines of the planar segments will also tend to be poor in areas where other objects occlude parts of the roof planes [Alharty and Bethel, 2004].

The goal of this chapter is twofold. First, we want to describe a method for roof plane delineation that eliminates user-defined thresholds as far as possible. This is achieved by making all

decisions dependent on statistical reasoning, relying on the framework of uncertain projective geometry [Heuel, 2004] and on robust estimation [Kager, 1989]. Second, we want to describe a new algorithm for the detection of step edges for delineating roof polygons, taking into account domain specific information in order to eliminate disturbances caused by trees adjacent to buildings.

7.1.2 Related Work

[Vosselman, 1999] described a method for the reconstruction of buildings by polyhedral models from LIDAR data. His algorithm for planar segmentation operates on a Delaunay triangulation of the original LIDAR points. The initial roof boundaries are given by the edges of the outmost triangles of the roof planes. Two planes are considered to intersect if the distance between their outlines is small. Step edges are assumed to be either parallel or orthogonal to the main direction of the building, and a merging algorithm is used to obtain sequences of boundary points belonging to the same straight line segment. By these assumptions, the algorithm is restricted to buildings only having right-angled corners at their outlines. [Vögtle and Steinle, 2000] determine step edges as roof plane outlines of significant elevation difference and estimate the position of these edges by an adjustment procedure taking into account the maximum gradient of the DSM in the vicinity of the edge. The authors do not describe how they determine the shapes of more complex step edges. Relying on the maximum gradient of the DSM is critical at building outlines. This is also acknowledged by [Alharty and Bethel, 2004]. They thin out the initial roof boundary polygons derived from the outlines of the planes by subsequently eliminating points that are determined not to contribute significantly to the polygon. This results in isolated polygons for each roof plane, which are not necessarily connected. Neighbouring polygon segments are aligned if their 2D distance is below a threshold, and vertices are merged if their 3D distance is small. No adjustment of the vertices is carried out apart from computing their average position.

[Sze et al., 1998] and [Jiang and Bunke, 1999] described edge extraction algorithms from range images. These algorithms detect edges at discontinuities of both height and slope. In the context of building reconstruction, step edges extracted by a generic edge extractor would have to be matched with the approximate roof outlines. In the case of trees adjacent to buildings, the edge extractor is likely to determine the outline of the trees rather than the building outline. In [Rottensteiner, 2003], it was shown how planes can be detected in a LIDAR DSM. Edge pixel candidates were determined at positions of maximum height gradient. Again, this resulted in problems where trees were adjacent to buildings. In order to overcome such problems, we propose to use a specific step edge extraction technique that takes into account domain specific information for detecting edge candidate pixels.

7.2 Background

7.2.1 Workflow for Automated Building Reconstruction

In this work, we assume the locations of buildings to be known. We use the algorithm described in [Rottensteiner et al., 2005b] for building detection. The building outlines are only known with an accuracy of up to 1 - 3 m. The accuracy of a LIDAR point is given by the standard deviations σ_P and σ_Z of a planimetric co-ordinate and the height. We chose $\sigma_P = \pm 25$ cm and $\sigma_Z = \pm 7.5$ cm. First, the LIDAR data are sampled into a DSM in the form of a regular grid of width Δ by linear prediction. For

the examples in this chapter, the point spacing of the LIDAR data was 1.2 m, and we chose Δ = 0.5 m. The work flow for the geometric reconstruction of buildings consists of four steps:

1. **Detection of roof planes** based on a segmentation of the DSM to find planes which are expanded by region growing.
2. **Grouping of roof planes and roof plane delineation:** Co-planar roof segments are merged, and hypotheses for intersection lines and/or step edges are created based on an analysis of the neighbourhood relations.
3. **Consistent estimation of the building parameters** to improve these parameters using all available sensor data.
4. **Model regularisation** by introducing hypotheses about geometric constraints into the estimation process.

In this chapter we will focus on the second stage, describing a new algorithm for step edge detection and showing how statistical tests and robust estimation can be applied to make decisions in the reconstruction process.

7.2.2 Representation of Geometric Entities

In this work, we represent geometric entities by their homogeneous co-ordinates and by the variance-covariance matrices of these co-ordinates [Heuel, 2004]. Each vector consists of a homogeneous part and Euclidean part; a Euclidean representation (which is essentially required for graphical output) can be achieved by dividing the vector by the norm of the homogeneous part. The variance-covariance matrix of the Euclidean representation can be derived by error propagation. To avoid numerical problems, the centre of the co-ordinate system is shifted to the centre of the building.

- With 2D and 3D points, we generally use the Euclidean representation: $\mathbf{X}^{2D} = (X, Y, 1)^T$ and $\mathbf{X}^{3D} = (X, Y, Z, 1)^T$. For the variance-covariance matrix $\mathbf{Q_X}$ of the LIDAR points we assume $q_{XY} = q_{XZ} = q_{YZ} = 0$, $q_{XX} = q_{YY} = \sigma_P^2$ and $q_{ZZ} = \sigma_Z^2$.
- 2D lines are represented by $\mathbf{L}^{2D} = (A, B, W)^T$, where the homogeneous part $(A, B)^T$ is the normal vector of the line. The rank of the variance-covariance matrix $\mathbf{Q}_{L2D}$ is 2. 2D edge segments, i.e. polygon segments at step edges, are represented by $\mathbf{L}^{2D}$, $\mathbf{Q}_{L2D}$, their endpoints $\mathbf{X}_1^{2D}$ and $\mathbf{X}_2^{2D}$, and their centre point $\mathbf{X}_C^{2D}$. $\mathbf{L}^{2D}$ and $\mathbf{Q}_{L2D}$ are estimated from the edge candidate points assigned to the edge segments by minimising the squared sum of the distances of these points from the line.
- 3D planes are represented by vectors $\mathbf{P} = (A, B, C, W)^T$, where the homogeneous part $(A, B, C)^T$ is the normal vector of the plane. The rank of the variance-covariance matrix $\mathbf{Q_P}$ is 3. We also store the centre point $\mathbf{X}_C^{3D}$ of the plane. $\mathbf{P}$ and $\mathbf{Q}_P$ are estimated from the DSM points assigned to the plane by minimising the squared sum of the distances of these points from the plane.
- 3D lines are represented by vectors $\mathbf{L}^{3D} = (L_1, L_2, L_3, L_4, L_5, L_6)^T$, where the homogeneous part $(L_1, L_2, L_3)^T$ is the directional vector of the line. If the line is constructed from two points

$\mathbf{X}_1^{3D}$ and $\mathbf{X}_2^{3D}$, the vector $(L_4, L_5, L_6)^T$ can be interpreted as the cross-product of $\mathbf{X}_1^{3D}$ and $\mathbf{X}_2^{3D}$. The rank of the variance-covariance matrix $\mathbf{Q}_{L3D}$ is 4. 3D edge segments are represented by $\mathbf{L}^{3D}$, $\mathbf{Q}_{L3D}$, and their endpoints $\mathbf{X}_1^{3D}$ and $\mathbf{X}_2^{3D}$. $\mathbf{L}^{3D}$ and $\mathbf{Q}_{L3D}$ are derived by the intersection of two planes.

7.2.3 Testing of Geometric Relations

In order to test whether two geometric entities $\mathbf{N}$ and $\mathbf{M}$ fulfil a certain geometric relation, a distance metric $\mathbf{d}_{NM}$ and its variance-covariance matrix $\mathbf{Q}_{NM}$ can be computed:

$$\begin{aligned} \mathbf{d}_{NM} = & \quad \mathbf{A}(\mathbf{N}) \cdot \mathbf{M} = \mathbf{B}(\mathbf{M}) \cdot \mathbf{N} \\ \mathbf{Q}_{NM} = & \quad \mathbf{A}(\mathbf{N}) \cdot \mathbf{Q}_M \cdot \mathbf{A}^T(\mathbf{N}) + \mathbf{B}(\mathbf{M}) \cdot \mathbf{Q}_N \cdot \mathbf{B}^T(\mathbf{M}) \end{aligned} \tag{7.1}$$

In equation 7.1, $\mathbf{A}(\mathbf{N})$ and $\mathbf{B}(\mathbf{M})$ are matrices depending on $\mathbf{N}$ and $\mathbf{M}$, respectively. Table 7.1 sums up the definitions of $\mathbf{A}$ and $\mathbf{B}$ for the relations that are of interest in the context of our work. The definitions of the construction matrices used in table 7.1 are given by equations 7.2 [Heuel, 2004].

Relation	$\mathbf{N}$	$\mathbf{M}$	$\mathbf{A}(\mathbf{N})$	$\mathbf{B}(\mathbf{M})$	dof
Identity	$\mathbf{L}_1^{2D}$	$\mathbf{L}_2^{2D}$	$\mathbf{S}\left(\mathbf{L}_1^{2D}\right)$	$-\mathbf{S}\left(\mathbf{L}_2^{2D}\right)$	2
Incidence	$\mathbf{X}^{3D}$	$\mathbf{P}$	$\left(\mathbf{X}^{3D}\right)^T$	$\mathbf{P}^T$	1
Incidence	$\mathbf{L}^{3D}$	$\mathbf{X}^{3D}$	$\overline{\mathbf{\Gamma}}^T\left(\mathbf{L}^{3D}\right)$	$\overline{\mathbf{\Pi}}^T\left(\mathbf{X}^{3D}\right)$	2

Table 7.1: Definitions of the matrices $\mathbf{A}$ and $\mathbf{B}$ in equation 7.1 [Heuel, 2004]. dof: Degrees of freedom of the test.

$$\begin{aligned} \overline{\mathbf{\Pi}}\left(\mathbf{X}^{3D}\right) = & \begin{pmatrix} 0 & -Z & Y & 0 \\ Z & 0 & -X & 0 \\ -Y & X & 0 & 0 \\ 1 & 0 & 0 & -X \\ 0 & 1 & 0 & -Y \\ 0 & 0 & 1 & -Z \end{pmatrix} \\ \mathbf{S}\left(\mathbf{X}^{2D}\right) = & \begin{pmatrix} 0 & -1 & Y \\ 1 & 0 & -X \\ -Y & X & 0 \end{pmatrix} \\ \overline{\mathbf{\Gamma}}\left(\mathbf{L}^{3D}\right) = & \begin{pmatrix} 0 & L_3 & -L_2 & -L_4 \\ -L_3 & 0 & L_1 & -L_5 \\ L_2 & -L_1 & 0 & -L_6 \\ L_4 & L_5 & L_6 & 0 \end{pmatrix} \end{aligned} \tag{7.2}$$

The dimension of the distance vector $\mathbf{d}_{NM}$ according to table 7.1 and equations 7.1 and 7.2 might be higher than the degree of freedom (dof) of the test, resulting in a singular variance-covariance matrix $\mathbf{Q}_{NM}$. Thus, dof rows of $\mathbf{A}(\mathbf{N})$ and $\mathbf{B}(\mathbf{M})$ have to be selected in order to obtain a reduced distance vector $\mathbf{d}'_{NM}$:

$$\begin{aligned} \mathbf{d}'_{NM} = & \quad \mathbf{A}'(\mathbf{N}) \cdot \mathbf{M} = \mathbf{B}'(\mathbf{M}) \cdot \mathbf{N} \\ \mathbf{Q}'_{NM} = & \quad \mathbf{A}'(\mathbf{N}) \cdot \mathbf{Q}_M \cdot \mathbf{A}'^T(\mathbf{N}) + \mathbf{B}'(\mathbf{M}) \cdot \mathbf{Q}_N \cdot \mathbf{B}'^T(\mathbf{M}) \end{aligned} \tag{7.3}$$

$\mathbf{A}'$ and $\mathbf{B}'$ are the reduced matrices. From $\mathbf{d}'_{NM}$, a quantity t_{NM} following a χ^2_{dof} distribution can be derived:

$$t_{NM} = \mathbf{d}'^T_{NM} \cdot \mathbf{Q}'^{-1}_{NM} \cdot \mathbf{d}'_{NM} \tag{7.4}$$

Using a significance level α, a hypothesis is accepted if t_{NM} is smaller than the $(1-\alpha)$ - quantile $\chi^2_{1-\alpha;dof}$ of the χ^2_{dof} distribution with dof degrees of freedom. Hypotheses H_i about geometrical relations can be ranked according to the ratios between the test statistics t^i_{MN} and the quantiles $\chi^2_{1-\alpha;dof(i)}$.

7.3 Roof plane delineation

7.3.1 Detection of Roof Planes

For roof plane detection we use the iterative scheme of seed region detection and region growing described in [Rottensteiner, 2003]. In region growing, each point $\mathbf{X}^{3D}$ of the DSM adjacent to the seed region has to be tested whether or not it belongs to the plane $\mathbf{P}$. In order to speed up the computation, the variance σ_d^2 of the distance of $\mathbf{X}^{3D}$ from $\mathbf{P}$ is computed only once for a fixed point at a certain distance from $\mathbf{P}$'s centre point $\mathbf{X}^{3D}_C$, so that in the region growing process the distance of each point has to be compared to a fixed threshold d_{max}:

$$d_{max} = \sqrt{\chi^2_{1-\alpha;1}} \cdot \sigma_d \tag{7.5}$$

This is justified by the fact that σ_d^2 is dominated by the uncertainty of $\mathbf{X}^{3D}$ and the Euclidean part of $\mathbf{P}$. After each iteration, the roof plane parameters are recomputed and a decision is taken on whether or not a detected segment actually corresponds to a plane. This is done by comparing the r.m.s. error σ_0 of unit weight of the planar adjustment to a user-defined threshold σ_{Pmax}. It could, however, be replaced by a test of σ_0, comparing it to the accuracy of a LIDAR point. Here, we choose $\sigma_{Pmax} = 2 \cdot \sigma_Z$.

Our iterative scheme of seed region selection and region growing yields an oversegmentation of the DSM [Rottensteiner, 2003]. Co-planar neighbouring planes have to be merged after segmentation. For each pair of neighbouring planes $\mathbf{P}_i$ and $\mathbf{P}_j$ we compute the parameters of the combined plane as well as its r.m.s. error of unit weight σ_c. The ratio $F = \sigma_c^2/\sigma_s^2$ between σ_c and the r.m.s. error of unit weight σ_s of a separate adjustment of the two planes follows a Fisher distribution with $f_c = n_i + n_j - 3$ and $f_s = n_i + n_j - 6$ degrees of freedom, where n_i and n_j are the numbers of DSM points assigned to $\mathbf{P}_i$ and $\mathbf{P}_j$, respectively. In order to compare hypotheses about the co-planarity of planes, we compute the ratio $r_{ij} = F/F_{fc,fs,1-\alpha} = \sigma_c^2/(\sigma_s^2 \cdot F_{fc,fs,1-\alpha})$, where $F_{fc,fs,1-\alpha}$ is the $(1-\alpha)$ - quantile of the Fisher distribution. Two planes $\mathbf{P}_i$ and $\mathbf{P}_j$ are considered to be co-planar if $r_{ij} < 1$. As this turned out to be too pessimistic, we introduced a second criterion, accepting planes to be co-planar if σ_c is below a certain threshold. All co-planar pairs $\mathbf{P}_i$ and $\mathbf{P}_j$ are ranked according to

r_{ij}. The pair receiving the minimum value of r_{ij} is merged, and the co-planarity ratios r_{ij} are recomputed for the remaining planes. This process is repeated until no further planes can be merged. The upper part of figure 7.1 shows the planar segmentation for a simple roof partly occluded by trees, whilst the lower part shows a more complex industrial building.

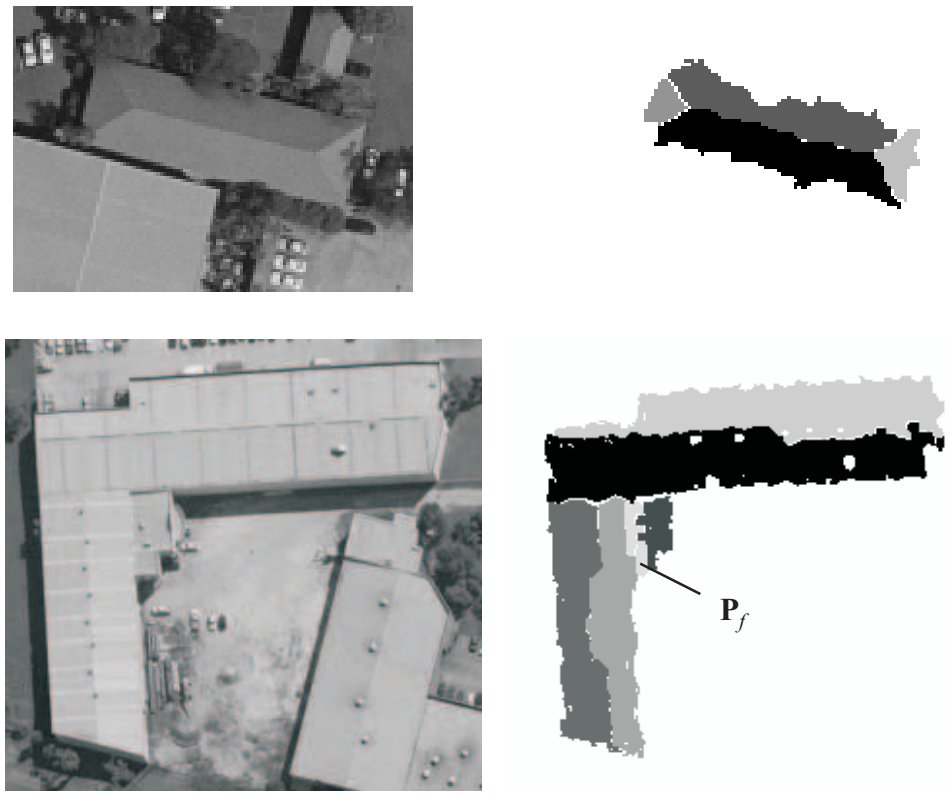


Figure 7.1: Orthophoto (left) and planar segments (right) for two buildings. Width of upper window: 60 m; lower window: 115 m. Plane $\mathbf{P}_f$ will be eliminated later.

7.3.2 Classification of Neighbourhood Relations

Once the roof planes have been detected, their boundary polygons are determined. We create a Voronoi diagram of the planar segments. The boundaries of the planes $\mathbf{P}_i$ in the Voronoi diagram deliver approximate values for the boundary polygons p_i of these planes. By using the Voronoi diagram for approximations, we overcome segmentation problems such as gaps between neighbouring planes (e.g. caused by chimneys), or incomplete planes due to occluding trees. The heights of the vertices of p_i are computed using the parameters of the plane $\mathbf{P}_i$. The polygons p_i are split into an ordered set of polygon segments $p_{i,j,k}$, where each segment $p_{i,j,k}$ separates plane $\mathbf{P}_i$ from its neighbouring plane $\mathbf{P}_j$. (The index k refers to the sequential position of $p_{i,j,k}$ within p_i, whereas the index j denotes the neighbouring plane. Unlike j, k can thus only occur once in p_i). Then, each polygon segment $p_{i,j,k}$ is classified according to whether it corresponds to a step edge or to an intersection line. This classification has to take into account the uncertainty of the planes $\mathbf{P}_i$ and $\mathbf{P}_j$ and of the approximate positions of the vertices $\mathbf{X}^l_{i,j,k}$ of $p_{i,j,k}$. We test all vertices of $p_{i,j,k}$ whether they are co-incident with the intersection line $\mathbf{L}^{3D}$ of $\mathbf{P}_i$ and $\mathbf{P}_j$, in the way described in section 7.2.3. The

standard deviation of a planimetric co-ordinate σ_P of $\mathbf{X}^l_{i,j,k}$ has to reflect the fact that the boundaries of the Voronoi diagram are more uncertain for roofs having a small tilt δ. σ_P also depends on the distance d_i of the point from the nearest point actually assigned to the plane $\mathbf{P}_i$:

$$\sigma_P = \sqrt{\left[\sigma_Z \cdot cot\left(\delta\right)\right]^2 + d_i^2} \tag{7.6}$$

For small δ, we limit σ_P by a threshold σ_{Pmax}. If all vertices of $p_{i,j,k}$ are found to be incident with the intersection line $\mathbf{L}^{3D}$, $p_{i,j,k}$ is classified as an intersection. If $p_{i,j,k}$ is an outer boundary or if no vertex is found to be on $\mathbf{L}^{3D}$, $p_{i,j,k}$ is classified as a step edge. If some vertices of $p_{i,j,k}$ are determined to be on $\mathbf{L}^{3D}$ and others are not, $p_{i,j,k}$ will be split up into several new segments, each having a different classification. Of these new segments, any segment smaller than $2 \cdot \Delta$ and intersection segments whose average distance from the approximate polygon is larger than the segment length are discarded.

A further consistency check is carried out for intersection segments. The planimetric positions of the boundaries between neighbouring roof planes are very uncertain especially with flat roofs. In the lower building of figure 7.1, the roof planes have a tilt of about 2^o, and the initial boundaries are about 2.5 m from the intersection line. Replacing the original boundaries by the intersection line will cause many points originally assigned to plane $\mathbf{P}_i$ to be within the boundary of $\mathbf{P}_j$ and vice versa (figure 7.2). Therefore, we compute the planar parameters of the cut-off segments $\mathbf{P}_{ij}$ and $\mathbf{P}_{ji}$. If $\mathbf{P}_{ji}$ is co-planar with $\mathbf{P}_i \backslash \mathbf{P}_{ij}$ (i.e. $\mathbf{P}_i$ without the points belonging to $\mathbf{P}_{ij}$) and if $\mathbf{P}_{ij}$ is co-planar with $\mathbf{P}_j \backslash \mathbf{P}_{ji}$, then the hypothesis that $\mathbf{L}^{3D}$ is the boundary between $\mathbf{P}_i$ and $\mathbf{P}_j$ is accepted; otherwise, a step edge is assumed. The left part of figure 7.3 shows the results of the classification of the segments $p_{i,j,k}$ for the two buildings in figure 7.1. [1]

Figure 7.2: Two planes $\mathbf{P}_i$, $\mathbf{P}_j$. $\mathbf{P}_{ij}$ is cut off $\mathbf{P}_i$ by the intersection line $\mathbf{L}^{3D}$, and $\mathbf{P}_{ji}$ is the part of $\mathbf{P}_j$ cut off by $\mathbf{L}^{3D}$.

7.3.3 Detection of Step Edges

7.3.3.1 Detection of Candidate Points

Step edges correspond to the positions of maximum height changes. However, this is only true where no other objects interfere with the roof planes. That is why we include knowledge about the nature of roof planes into the extraction process. This process is different for step edges at outer boundaries and those separating two roof planes (figure 7.4). The original polygons $p_{i,j,k}$ are sampled at Δ. For each vertex of $p_{i,j,k}$, we try to determine one edge candidate point by analysing a profile of the DSM that is orthogonal to $p_{i,j,k}$ and passes through that vertex. The profiles should be long enough to make sure that they partly cover $\mathbf{P}_i$. All profiles are ordered from the interior of $\mathbf{P}_i$ to its exterior.

[1] The left part of the figure only contains intersections as final polygon segments (red), since the final positions of the step edges have not yet been determined.

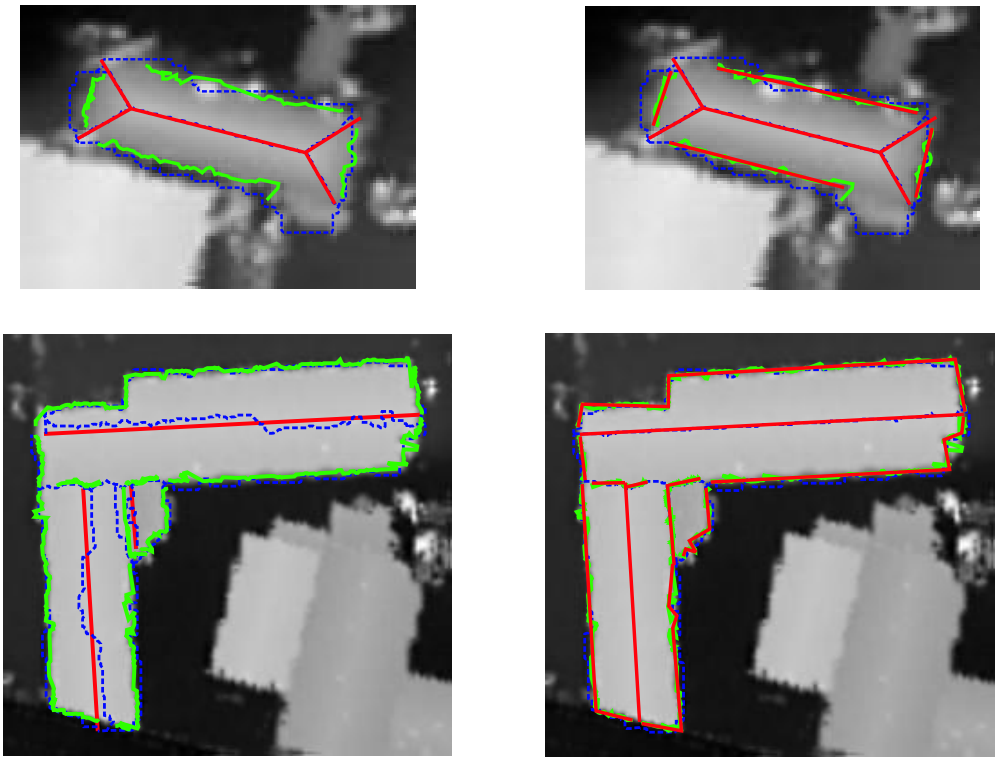


Figure 7.3: Blue dashed lines: approximate boundary polygons. Green: original step edges. Red: final polygon segments. Left: before generalisation of step edges. Right: after improving the planar segmentation. A colour version of this figure can be found on page 131.

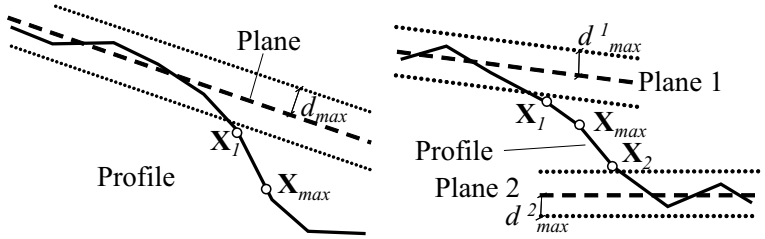


Figure 7.4: Left: Step edge detection at building outlines. $\mathbf{X}_1$: the first point on the profile outside the tolerance band of width d_{max}. $\mathbf{X}_{max}$: point of maximum height gradient. Right: a step edge between two planes. $\mathbf{X}_1$ and $\mathbf{X}_2$: the first points outside the tolerance bands.

At outer boundaries, we look for the first point on the profile that is not on the plane $\mathbf{P}_i$, i.e. for the first point $\mathbf{X}_1$ having a distance larger than d_{max} (cf. equation 7.5) from $\mathbf{P}_i$. If no such point is found, the profile is supposed not to contain the step edge. Otherwise, the point $\mathbf{X}_{max}$ of maximum height difference is searched for on the profile starting from $\mathbf{X}_1$. The height difference between neighbouring points must be negative, because the terrain has to be lower than the roof. The search for a maximum is thus stopped if the height difference becomes positive, which happens if the roof boundary is occluded by a high tree. A further criterion is that the point $\mathbf{X}_{max}$ must be below the roof plane, otherwise it is discarded. By this criterion we eliminate points on low vegetation next to a roof. Where no candidate point is found, the step edge is thus assumed to be a straight line between

the two closest step edge points visible to the sensor, which is exactly what a human operator would do in such a situation. For instance, in the upper building of figure 7.3, all corners are close to and partly occluded by trees. Building detection has included the group of trees in the right lower corner to the building. No incorrect edge points are determined.

With step edges separating two roof planes, we search for the first points $\mathbf{X}_1$ and $\mathbf{X}_2$ that are inconsistent with the two planes in a similar way. If the order of $\mathbf{X}_1$ and $\mathbf{X}_2$ is reversed, no step edge point can be determined; otherwise the position $\mathbf{X}_{max}$ of the step edge is determined as the position of the maximum height difference between $\mathbf{X}_1$ and $\mathbf{X}_2$.

7.3.3.2 Step Edge Generalisation

The detected edge polygons appear very noisy (figure 7.3) and need to be thinned out in a suitable manner. Again, we will mostly rely on statistical tests. However, we require one user-specified threshold which describes a degree of generalisation: the maximum length l_{max} of a polygon segment that can be discarded. We chose l_{max} to be 2 m and thus about two times the original point distance. First we eliminate points having a distance larger than l_{max} from both their predecessor and successor in the edge pixel chain. These outliers occur at profiles that belong to short segments of $p_{i,j,k}$ corresponding to noise. The remaining edge pixel chain is split into 2D segments $\mathbf{L}_n^{2D}$ by a simple recursive splitting algorithm. The parameters of $\mathbf{L}_n^{2D}$ are computed from an adjustment considering all edge points assigned to them. Segments containing less than three edge points are discarded. Then we test each pair of neighbouring 2D segments whether they are identical in the way described in section 7.2.3. We merge the pair possessing the best test statistic t_{L2DL2D} (equation 7.4) and recompute that test statistic for the neighbours of the merged segment. This procedure is repeated until no further segments can be merged. It turned out to be advantageous to replace the r.m.s. error of unit weight of the 2D segments by a fixed value equal to the original point spacing of the LIDAR data because the actual estimates derived in the adjustment were too optimistic. We also consider segments with a combined r.m.s. error of unit weight smaller than Δ to be candidates for merging. In a second merging step we search for segments $\mathbf{L}_n^{2D}$ shorter than l_{max}. For those segments, we check whether their neighbours $\mathbf{L}_{n-1}^{2D}$ and $\mathbf{L}_{n+1}^{2D}$ are identical. If this is the case, $\mathbf{L}_n^{2D}$ is eliminated, and $\mathbf{L}_{n-1}^{2D}$ and $\mathbf{L}_{n+1}^{2D}$ are merged. Finally, the vertices of the polygon segment corresponding to the step edge are determined by intersecting neighbouring 2D segments. If the segments are nearly parallel, replacing the segment endpoints by the intersection point might change the direction one of these segments. If this is the case, or if more than 30% of one segment were cut off by the intersection, the neighbouring end points are connected by a polygon edge.

7.3.3.3 Improving the Planar Segmentation

Now that the step edges have been determined precisely, the roof polygons p_i are obtained independently by a concatenation of their respective segments. The polygons are shifted with respect to their original positions. This affects the original segmentation and thus also the neighbourhood relations. That is why, after generating the roof polygons p_i, all pixels inside p_i assigned to a plane other than $\mathbf{P}_i$ and all pixels assigned to $\mathbf{P}_i$ outside p_i are eliminated from their respective planes. This is followed by a new iteration of region growing, where first the regions are only allowed to grow within their polygons p_i. Thus an improved segmentation is obtained. Small segments such as the plane $\mathbf{P}_f$ in figure 7.1 might be eliminated in this process. Having improved the segmentation, the

boundary classification and step edge detection are repeated. The right part of figure 7.3 shows the final positions of all segments $p_{i,j,k}$ for the buildings in figure 7.1.

7.3.4 Combination of Roof Polygon Sections

Until now, all polygon segments $p_{i,j,k}$ were handled individually. Therefore, having determined the positions of these segments, consistency checks have to be carried out. First, the internal consistency of each polygon p_i is checked. If there are two consecutive segments $p_{i,j,k}$ and $p_{i,l,k+1}$ classified as intersections, we have to check whether the corresponding intersection lines themselves intersect. If replacing the segment endpoints $\mathbf{X}_{2,k}$ and $\mathbf{X}_{1,k+1}$ by the intersection point $\mathbf{I}$ changes the direction of one of the segments, a new step edge has to be inserted between $\mathbf{X}_{2,k}$ and $\mathbf{X}_{1,k+1}$ (figure 7.5). Second, we have to ensure that for each segment $p_{i,j,k}$ belonging to the boundary of plane $\mathbf{P}_i$, but not to the building outline, there is a matching opposite segment $p_{j,i,l}$ of the same type belonging to plane $\mathbf{P}_j$. If no such segment is found, it has to be inserted. With step edges, the 2D segments making up $p_{i,j,k}$ and $p_{j,i,l}$ have to be matched. We evaluate two measures between two 2D segments $\mathbf{L}_m^{2D} \in p_{i,j,k}$ and $\mathbf{L}_n^{2D} \in p_{j,i,l}$: the test statistic t_{mn} (equation 7.4) and the overlaps between the segments, i.e. the lengths of the projections of $\mathbf{L}_m^{2D}$ to $\mathbf{L}_n^{2D}$ and vice versa. We determine that $\mathbf{L}_m^{2D}$ and $\mathbf{L}_n^{2D}$ match if the overlap is more than 50% for one of the segments and if a statistical test shows the 2D lines to be identical. Edge pixel chains of matching segments are merged. Segments without a match are discarded.

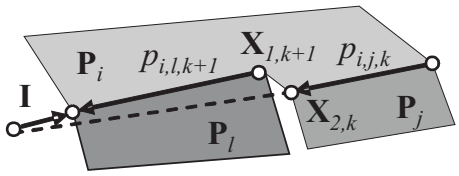


Figure 7.5: Checking consistency of $p_{i,j,k}$ and $p_{i,l,k+1}$.

Having obtained consistency of the polygon segments both within their polygons and with respect to their neighbouring planes, the polygon segments have to be combined. This involves an adjustment of the vertices $\mathbf{X}$ at the transitions between consecutive polygon segments $p_{i,j,k}$ and $p_{i,j,k+1}$. First, all polygon segments intersecting at one planimetric position have to be found. Figure 7.6 shows an example involving three planes of which two intersect. There are altogether three polygon segments. One of them is the intersection line $\mathbf{L}^{3D}$, and the other two are step edges. Of the polygon segments corresponding to the step edges, the two 2D segments $\mathbf{L}_{13}^{2D}$ and $\mathbf{L}_{23}^{2D}$ closest to the intersection point are considered. There are two intersection points $\mathbf{X}_1$ and $\mathbf{X}_2$ having the same planimetric position but different heights. For adjustment, we consider all the planes in the vicinity of the vertices $\mathbf{X}_1$ and $\mathbf{X}_2$, i.e. the roof planes ($\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$) and the walls corresponding to the step edge segments ($\mathbf{L}_{13}^{2D}$, $\mathbf{L}_{23}^{2D}$). For each plane, we observe the distance between the plane and the point $\mathbf{X}_i = (X, Y, Z_i)^T$ to be 0. In order to model the uncertainty of the planes and step edges, we introduce approximate values $\mathbf{X}_i^0 = (X^0, Y^0, Z_i^0)^T$ and compute the weights of the distance observations from the standard deviations σ_i of the distances between the approximate position $\mathbf{X}_i^0$ and the respective plane. The observation equations for a roof plane u giving support to height Z_i and for a wall w look as follows:

$$
\begin{aligned}
0 + v_{ui} = &\quad A_u \cdot (X^0 + \delta X) + B_u \cdot (Y^0 + \delta Y) + C_u \cdot (Z_i^0 + \delta Z_i) + W_u \\
0 + v_w = &\quad A_w \cdot (X^0 + \delta X) + B_w \cdot (Y^0 + \delta Y) + W_w \\
\sigma_{ui}^2 = &\quad \mathbf{X}_i^{0T} \cdot \mathbf{Q}_{Pu} \cdot \mathbf{X}_i^0 \\
\sigma_w^2 = &\quad (X^0, Y^0) \cdot \mathbf{Q}_{Lw} \cdot (X^0, Y^0)^T
\end{aligned}
\tag{7.7}
$$

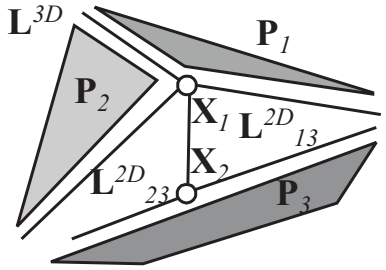


Figure 7.6: Vertex adjustment.

In equations 7.7, v denotes the correction of the observation. Equations 7.7 are used in an iterative least squares adjustment. After each iteration, the approximate co-ordinates are improved by the estimates $(\delta X, \delta Y, \delta Z_i)$, and the weights are re-computed. This model assumes that all walls intersect in one vertical line. Due to errors in step edge extraction, small step edge segments might have been missed and the extracted step edge segment might pass by the intersection point at (X, Y). To find such segments, we compute the normalised corrections $v_w^n = v_w/\sigma_w$ of the wall observations after each iteration and exclude the wall with a maximum value of v_w^n if $v_w^n > 3.5$. For all excluded walls, a new step edge segment is introduced between the original end point of the step edge and the adjusted position of the vertex. The left part of figure 7.7 shows the resulting roof boundaries.

7.3.5 Regularisation and Adjustment

After the vertices have been adjusted, a consistent polyhedral model in boundary representation is created. Only local information was used for the adjustment of the vertices. In an overall adjustment, all observations (original LIDAR points, 2D positions of step edges) should be used to determine the parameters of all planes and vertices simultaneously. The model can then be checked for geometrical regularities. Where evidence for such regularities is found, they can be considered in a final adjustment. In [Rottensteiner, 2003] we have presented the adjustment model and the way such regularities can be considered. The adjustment module has been implemented but has not yet been integrated into the algorithm. The "constraint generator" checking for geometric regularities has not yet been implemented. The right part of figure 7.7 shows the adjusted roof polygons with manually added constraints back-projected to an areal image of the area.

7.4 Results and Discussion

Figures 7.7 and 7.8 show some buildings extracted from the LIDAR data in our test data set. In general, the roof structure is well preserved in the models. Corners are not forced to be right-angled. The

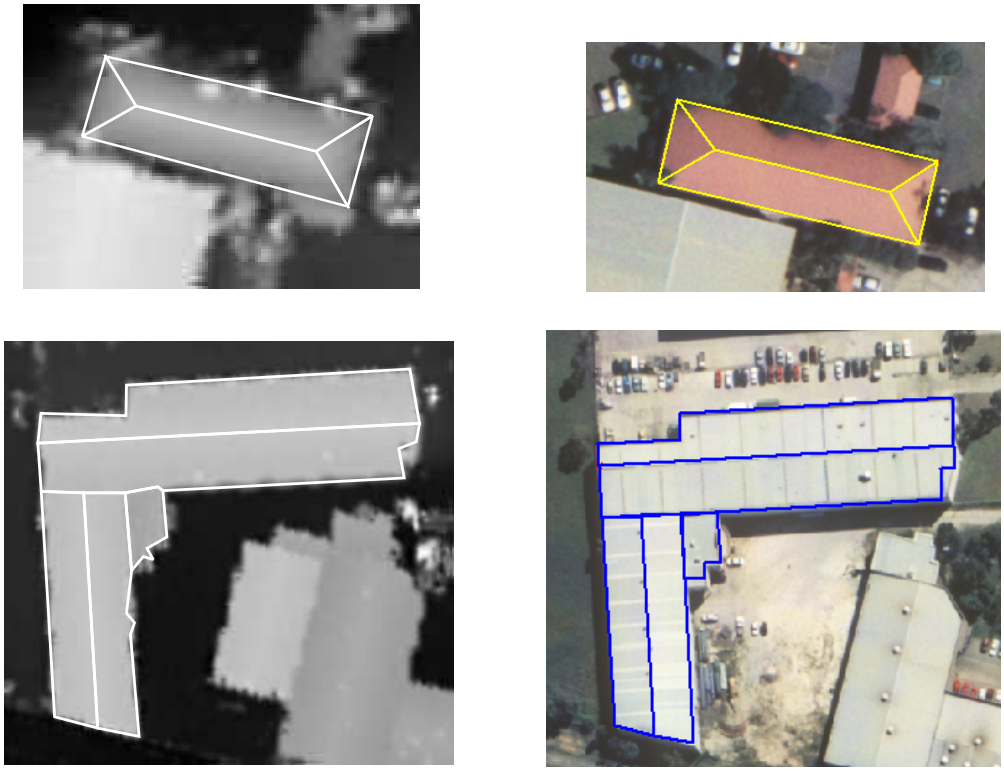
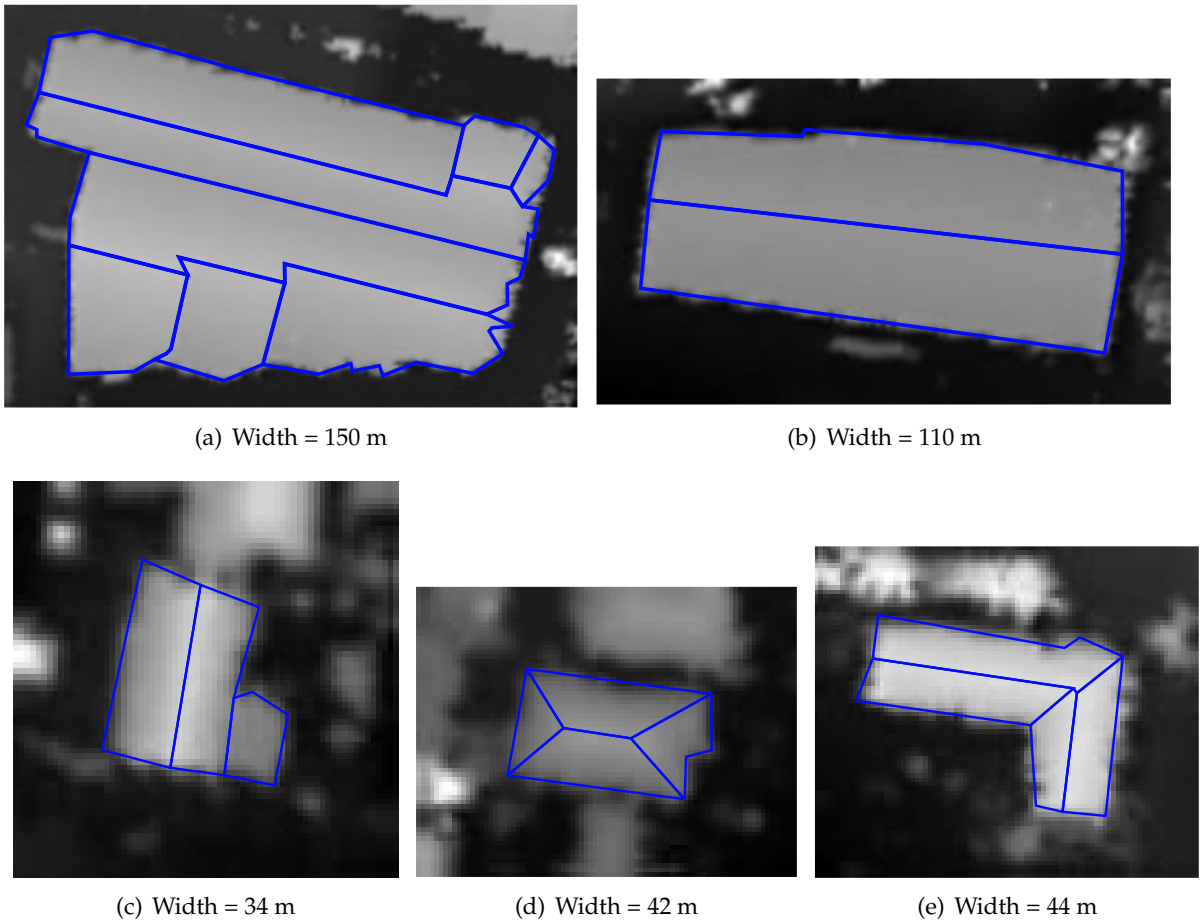


Figure 7.7: Left: roof polygons after adjustment of vertices. Right: after overall adjustment with constraints (colours selected for visibility).

outline of the buildings still looks a bit ragged, and in some cases as in figure 7.8(c) the regularisation would improve the extracted model considerably. Figure 7.7d shows the limits of the plane extraction method: even though the general roof shape has been captured, the rightmost plane actually merges a smaller protruding building part consisting of two planes; the larger of these planes has six LIDAR hits, the smaller one has none. The industrial (lower) building in figure 7.7 is in general correctly reconstructed; however, an outlier in the LIDAR data caused a "hole" in the DSM which resulted in one of the corners on the right edge of the building being cut off. Figure 7.7 shows how the visual appearance is improved by the regularisation and the overall adjustment. The basis for such an adjustment is a topologically correct model, which can be systematically scanned for geometrical regularities. The industrial buildings in our examples (lower part of figure 7.7, figures 7.8(a) and 7.8(b)) are not easily reconstructed by primitives having a rectangular footprint. The r.m.s. error of planar fit is in the range of ± 5 cm to ± 10 cm for all planes. The internal precision of the adjusted building vertices in figure 7.7 is in the range of ± 5 cm to ± 30 cm in planimetry (depending on the intersection geometry, the geometric constraints, and the number of detected edge pixels for step edges) and ± 2 cm to ± 3 cm in height.

7.5 Conclusions

We have presented a method for roof plane delineation from LIDAR data that aims at the reconstruction of buildings by polyhedral models. No 2D GIS data were required, and our examples show



(a) Width = 150 m (b) Width = 110 m

(c) Width = 34 m (d) Width = 42 m (e) Width = 44 m

Figure 7.8: Five buildings extracted from the LIDAR data.

that models of a high level of detail can be reconstructed. Our method includes a new approach to step edge detection which should make the determination of building outlines less vulnerable to effects of adjoining trees. In the reconstruction process, most decisions are taken based on statistical tests or robust estimation, which is an important step towards making the reconstruction of buildings from LIDAR data more robust. The resulting building models are topologically correct if not yet regularised. Preliminary results using geometric constraints generated manually, based on visual inspection, show that using our method it is feasible to generate high-quality building models from LIDAR data alone.

Acknowledgements

This research was funded by ARC Linkage Project LP0230563 and ARC Discovery Project DP0344678. The Fairfield data set was provided by AAMHatch (http://www.aamhatch.com.au)

Chapter 8

Consistent Estimation of Building Parameters Considering Geometric Regularities by Soft Constraints

8.1 Introduction

The shapes of buildings and other man-made objects, despite being very complex in realistic scenes, are often characterised by certain geometrical regularities. At a level of detail typical for topographic mapping (mapping scales 1:500 to 1:10000) most buildings can be modelled by polyhedrons. This implies that all vertices belonging to a face must be situated on a plane in object space. Apart from that, other geometrical regularities include perpendicular walls, horizontal roof edges, or symmetry between roof faces.

It is the goal of automatic building reconstruction to generate 3D building models from sensor data in previously detected regions of interest. In this context, model regularisation by considering geometric constraints is essential for achieving high quality building models. Besides resulting in a more regular visual appearance, considering geometric regularities helps to improve the geometric accuracy of the models, especially if the sensor geometry is weak. There are two general strategies for building reconstruction, differing in the way buildings are represented in the reconstruction process and thus also in the way geometric regularities are considered. The first strategy is based on a bottom-up process. The sensor data are segmented in order to obtain 3D features such as edges and planes. These features are combined to obtain a polyhedral model, e.g. [Rottensteiner et al., 2005a]. Alternatively, buildings can be reconstructed by parametric primitives in a top-down process, e.g. [Brenner, 2000]. In the first case, assumptions on geometric regularities may or may not be used in order to select the 3D features and group them; they can and should be considered as additional information in a final parameter estimation process yielding consistent and regularised building models. In the second case, assumptions about regularities, e.g. rectangular footprints, are an implicit part of the description of the primitives. Using parametric primitives reduces the level of detail that can be achieved as the number of primitives is usually small and most have a rectangular footprint. This can be avoided by using "adaptive primitives" [Rottensteiner and Schulze, 2003], i.e. primitives having an adaptive parameterisation. However, the bottom-up strategy seems to be more flexible with respect to handling geometric regularities. They are not an implicit part of the building model, but rather are added as additional information to the

estimation of the building parameters and thus only have to be considered where enough evidence is found in the data. From the point of view of parameter estimation, this can be handled in two ways. First, geometric regularities can be considered in the adjustment by constraint equations. This strategy will result in models precisely fulfilling these "hard" constraints. Brenner [Brenner, 2005] has given an overview about the ways such constraints can be handled in object modelling. The alternative is to add "soft constraints", i.e. direct observations for entities describing a geometric regularity, to the adjustment of the sensor-based observations. In this case, the constraints will not be fulfilled exactly, but there will be residuals to the observations. The degree to which the constraints are fulfilled depends on the stochastic model. Using the second strategy, robust estimation techniques can be applied to the soft constraints to determine whether a hypothesis about a geometric regularity fits to the sensor data or not.

Vosselman [Vosselman, 1999] proposed an algorithm for building reconstruction from airborne laser scanner (ALS) data that determined building outlines under the assumption of all neighbouring walls intersecting at right angles. He addressed the necessity of adding constraints to the estimation of the model parameters without doing so himself. Vögtle and Steinle [Vögtle and Steinle, 2000] reconstruct buildings from ALS and spectral data. The coordinates of their building vertices are estimated by local adjustment only, and no geometric regularities are considered. Alharty and Bethel [Alharty and Bethel, 2004] describe two methods for roof outline detection. The first method relies on the existence of a dominant roof direction and the neighbouring walls being orthogonal. The second does not require such assumptions, but no overall adjustment is carried out, and no geometric regularities are considered. Ameri [Ameri, 2000b] describes a general adjustment model for building reconstruction from image data. Geometric constraints are considered. For instance, for two orthogonal building edges a direct observation of the inner product of the directional vectors is introduced. The weighting of such an algebraic observation seems to be somewhat critical. A method for fitting building models to multiple aerial images using "hard" constraints was presented in [Vallet and Taillandier, 2005]. [McGlone, 1996] describes the mathematical basis for handling geometrical constraints both as ("hard") condition equations and as "soft" constraints, using this basis for improving the results of multiple-image point matching under the assumption of certain object regularities.

In [Rottensteiner et al., 2005a] we have presented a method for automatic building reconstruction from ALS data that is based on the detection and combination of roof planes. The final step of building reconstruction is an overall adjustment of all observations to determine the model parameters consistently. The adjustment model was originally presented in [Rottensteiner, 2003], but implemented only recently. It is the first goal of this chapter to present this adjustment model in its improved and revised form and to show how it can be used as a tool for consistent estimation of building parameters for different types of available sensor data. Special emphasis is laid on the way geometric regularities can be considered. The second goal of this chapter is to evaluate the results of building reconstruction from ALS data by comparing automatically derived building models to reference data. This comparison should also show how effective the overall adjustment is in improving the geometric quality of building models.

8.2 Workflow for Building Reconstruction

Our algorithm for building reconstruction requires ALS points and a coarse approximation of the building outlines. The ALS data are sampled into a Digital Surface Model (DSM) in the

form of a regular grid of width Δ by linear prediction. The work flow consists of three steps [Rottensteiner et al., 2005a]:

1. **Detection of roof planes** based on a segmentation of the DSM. These planes are expanded by region growing.
2. **Grouping of roof planes and roof plane delineation:** Co-planar roof planes are merged, and hypotheses for intersection lines and/or step edges are created based on an analysis of the neighbourhood relations of the roof planes.
3. **Consistent estimation of the building parameters** to improve these parameters using all available sensor data and considering geometric constraints.

In step 2, the boundary polygons of the roof planes are determined as a combination of roof plane intersections and step edges, the step edges being located in the DSM by an edge extraction technique taking into account specific information about buildings. Decisions in the determination of the shapes of the roof polygons are based on hypothesis tests and robust estimation. We use the concept of uncertain projective geometry [Heuel, 2004] for consistent modelling of the stochastic properties of all geometric entities. In this chapter, we will focus on the final step of the reconstruction process.

8.3 The Adjustment Model

The adjustment problem we want to solve can be described as follows. We assume to have given a polyhedral building model in boundary representation (B-rep). The model consists of planar faces, loops, edges, and vertices. Each edge is the intersection of two neighbouring faces, and each vertex is the intersection of at least three planes of the model. All vertices belonging to the boundary of a face have to lie in the face's plane. The faces of the model are labelled as being a roof face, a wall, or the floor. Walls are modelled to be strictly vertical. The topology of the model and some meaningful initial values for its parameters are assumed to be known. The initial model can be the outcome of the bottom-up strategy for building reconstruction (cf. section 8.2). In this case it is an approximate version of the final model, and its initial parameters are already derived in some way from the sensor data. The coarse model has to be analysed for geometric regularities, which can be done automatically or based on the interaction of a human operator, and the model parameters have to be estimated. For that purpose, we use five categories of observations:

1. Observations representing the topology of the model
2. Observations corresponding to geometric regularities
3. Sensor and sensor-derived observations
4. Observations linking the sensor observations to the model
5. Direct observations for unknowns to avoid singularities.

They are used to determine four categories of unknowns:

1. The co-ordinates of the model vertices
2. The parameters of the model planes
3. Transformation parameters, e.g. the unknown angle for each pair of perpendicular walls (cf. section 8.3.2)
4. Additional unknowns, e.g. unknown object co-ordinates for each ALS point (cf. section 8.3.3.2).

Our method for handling the model topology and geometric regularities is independent not only from the types of sensor data that are used, but also from the way in which the original model was created. The adjustment model is based on the program ORIENT for hybrid photogrammetric adjustment, especially on its concept of handling object space constraints by "GESTALT" observations [Kager, 2000].

8.3.1 Observations Representing Model Topology

It is the idea of our method to find a mapping between the B-rep of the polyhedral model and a system of GESTALT observations representing the model topology in adjustment. GESTALT observations are observations of a point $\mathbf{P}$ being situated on a polynomial surface [Kager, 2000]. The polynomial is parameterised in an observation co-ordinate system (u, v, w) related to the object co-ordinate system by a shift $\mathbf{P}_0$ and three rotations $\mathbf{\Theta}^T = (\omega, \varphi, \kappa)^T$. The actual observation is $\mathbf{P}$'s distance from the surface which has to be 0. Using $(u_R, v_R, w_R)^T = \mathbf{R}^T(\mathbf{\Theta}) \cdot (\mathbf{P} - \mathbf{P}_0)$, with $\mathbf{R}^T(\mathbf{\Theta})$ being a transposed rotational matrix parameterised by $\mathbf{\Theta}$, and restricting ourselves to vertical planes for walls and tilted planes for roofs, there are three possible formulations of GESTALT observation equations:

$$r_u = \frac{m_u \cdot u_R + a_{00} + a_{01} \cdot m_v \cdot v_R}{\sqrt{1 + a_{01}^2}} \tag{8.1}$$

$$r_v = \frac{m_v \cdot v_R + b_{00} + b_{10} \cdot m_u \cdot u_R}{\sqrt{1 + b_{10}^2}} \tag{8.2}$$

$$r_w = \frac{m_w \cdot w_R + c_{00} + c_{10} \cdot m_u \cdot u_R + c_{01} \cdot m_v \cdot v_R}{\sqrt{1 + c_{10}^2 + c_{01}^2}} \tag{8.3}$$

In equations 8.1 - 8.3, r_i are the corrections of the fictitious observations of co-ordinate i and $m_i \in \{-1, 1\}$ are mirror coefficients. An application is free to decide which of the parameters ($\mathbf{P}$, $\mathbf{P_0}$, $\mathbf{\Theta}$, a_{jk}, b_{ik}, c_{ij}) are to be determined in adjustment and how to parameterise a surface. In addition, different GESTALTs can refer to identical transformation or surface parameters, which will be used to handle geometric regularities (cf. section 8.3.2). Here, we declare the rotations to be 0 and constant. $\mathbf{P_0}$ is a point situated inside the building and constant. For each face of the B-rep of the building model, we define a set of GESTALT observations, taking one of equations 8.1 and 8.2 for walls and equation 8.3 for roofs. The unknowns to be determined are the object co-ordinates of each vertex $\mathbf{P}$ and the plane parameters (a_{jk}, b_{ik}, c_{ij}). As each vertex is neighboured by at least three faces, the co-ordinates of the vertices are determined from these GESTALT observations and thus need not be observed directly in the sensor data. Further, these observations link the vertex co-ordinates to the

surface parameters and thus represent the building topology in the adjustment. They do already enforce geometric constraints by modelling walls as being strictly vertical and by declaring all vertices of a face to lie in the same plane. The stochastic model of these GESTALT observations is described by the a priori standard deviation σ_T of the fictitious distance between a point and the plane.

8.3.2 Observations Representing Geometric Regularities

Geometric regularities are considered by additional GESTALT equations, taking advantage of specific definitions of the observation co-ordinate system and specific parameterisations of the planes. Geometric regularities can occur between two planes or between two vertices of the model. In the current implementation, we restrict ourselves to regularities involving planes or vertices being neighbours of one edge. In all cases, the observation co-ordinate system is centred in one vertex $\mathbf{P}_1$ of that edge and the w-axis is vertical, thus $\omega = \varphi = 0 = const$. Four types of geometric regularities are considered (figure 8.1). The first type, a horizontal roof edge, involves the edge's end points: Its two vertices $\mathbf{P}_1$ and $\mathbf{P}_2$ must have identical heights. The two points are declared to be in a horizontal plane ε_h that is identical to the (u, v) - plane of the observation co-ordinate system. One observation is inserted for $\mathbf{P}_2$: $r_w = w_R = Z_2 - Z_1$.

The other cases involve the two neighbouring planes of an edge. One of the axes of the observation coordinate system is defined to be the intersection of these two planes ε_1 and ε_2. There is one additional unknown rotational angle κ describing the direction of the u-axis. For each vertex $\mathbf{P}_i$ of the planes, GESTALT observations are added for ε_1 or ε_2. For the edge's second vertex $\mathbf{P}_2$ two observations (one per plane) are added. The GESTALT observations for ε_1 and ε_2 are parameterised in a specific way:

- The edge is the intersection of two horizontal and symmetric roof planes ε_1 and ε_2. There is only one tilt parameter c_{01}^1. Symmetry is enforced by selecting $m_v = -1$ for ε_2:

$$\varepsilon_1 : r_w = \frac{w_R + c_{01}^1 \cdot v_R}{\sqrt{1 + \left(c_{01}^1\right)^2}}; \qquad \varepsilon_2 : r_w = \frac{w_R - c_{01}^1 \cdot v_R}{\sqrt{1 + \left(c_{01}^1\right)^2}} \tag{8.4}$$

- The edge is the intersection of two perpendicular walls: $\varepsilon_1 : r_u = u_R$, $\varepsilon_2 : r_v = v_R$. There is no additional surface parameter to be determined.

- Two walls are identical and the edge does not really exist in the object: $\varepsilon_1 : r_v = v_R$, $\varepsilon_2 : r_v = v_R$. There is no additional surface parameter. $\mathbf{P}_1$ and/or $\mathbf{P}_2$ might become undetermined, so that direct observations for one of the co-ordinates of these vertices have to be generated.

The stochastic model of these GESTALT observations is described by their a priori standard deviations σ_C. The "soft constraints" thus modelled will only be fulfilled up to a degree depending on σ_C. The GESTALT observations corresponding to the geometrical constraints can be subject to robust estimation for gross error detection. If the sensor observations contradict the constraints, the respective GESTALT observations should receive large residuals, which can be used to modulate the weights in an iterative robust estimation procedure [Kager, 1989]. Thus, if the GESTALT observations describing a geometric constraint are eliminated in adjustment, this means that the hypothesis about a constraint was wrong.

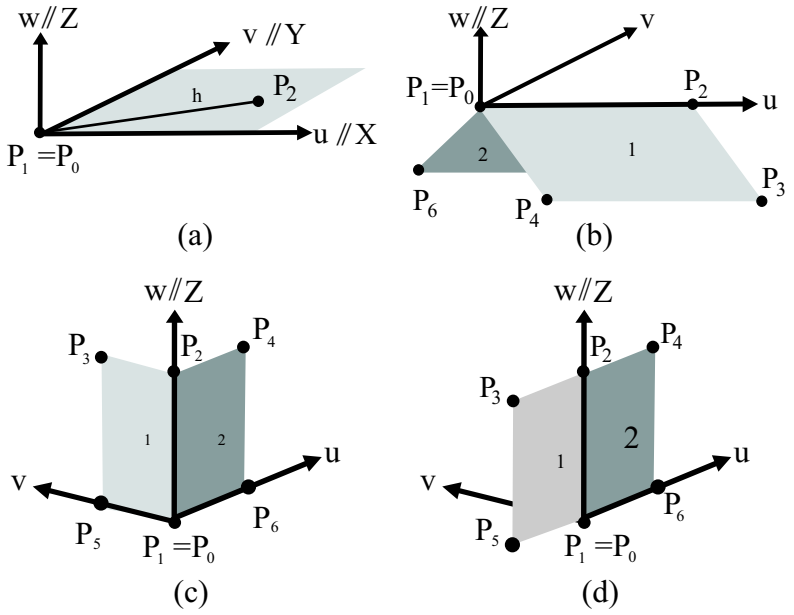


Figure 8.1: (a) Horizontal edge; (b) horizontal and symmetric edge; (c) perpendicular walls; (d) Identical walls.

Whether or not a hypothesis about a constraint is introduced can be decided in several ways. For instance, the coarse model can be analysed whether the angles between neighbouring walls differ from 90^o by less than a threshold ε_α, and a constraint about perpendicular walls can be inserted if this is the case. More sophisticated methods can take into account the stochastic properties of the coarse model. In a semi-automatic working environment, geometric constraints can be inserted (and enforced) by the user. The principle can be expanded to the definition of parametric primitives by generating more complex systems of constraints between the planes of a building [Rottensteiner and Schulze, 2003].

8.3.3 Sensor Observations and Observations Linking the Sensor Data to the Model

The observations described so far link the plane parameters to the vertices or to the parameters of other planes. In order to determine the surface parameters, observations derived from the sensor data are necessary. ORIENT can handle a large variety of sensor models. Any of these sensors or any combination of them can be used in adjustment. Here we will restrict ourselves to image and ALS data.

8.3.3.1 Image co-ordinates

Points measured in images are related to object space by the perspective equations. We assume the orientation parameters of the images to be known and constant. An observed image point has to be assigned to an entity of the object model to contribute to the determination of the model parameters. Two cases can be distinguished. First, an image point can be assigned to a building vertex, which yields two perspective observation equations for that vertex. Second, the image point can be assigned to a model edge. As such a point is not a part of the model, its object co-ordinates have

to be determined as additional unknowns; however, each point assigned to an object edge yields four additional observations: its two image co-ordinates and two GESTALT observations (one for each object plane intersecting at the object edge). The stochastic model of an image co-ordinate is described by its standard deviation σ_I. Depending on the way the image points were determined, σ_I can describe the accuracy of manual measurement, or it can be the result of a feature extraction process.

8.3.3.2 ALS data

ALS points give support to the determination of the roof plane parameters. As an ALS point is not a part of the model, its object co-ordinates have to be determined as unknowns. Each ALS point gives four observations, namely its three co-ordinates and one GESTALT observation for the roof plane the point is assigned to. As the walls only receive few laser hits, their parameters have to be determined from other observations. Walls correspond to sections of step edges in the DSM [Rottensteiner et al., 2005a]. Each step edge section is derived from "edge points" in the DSM (e.g. points of maximum height gradient). In order to determine the walls, these edge points have to be used as observations in a way similar to the original ALS points: Each edge point gives three observations (its X and Y co-ordinates and one GESTALT), but two additional unknowns (again X and Y). The ALS observations can be modelled in two different ways: They can be introduced as "control point" observations, i.e. as direct observations for the object co-ordinates, or they can be introduced as "model points". In the latter case, the ALS points are linked to the object co-ordinate system by a rigid motion, and the six parameters of that rigid motion are estimated in the adjustment. Using this variant, local shifts and rotations of the ALS co-ordinate system with respect to the object co-ordinate system that might be the result of systematic GPS and INS errors of the ALS system can be compensated. This only makes sense if additional data, e.g. aerial images, are available. Otherwise, the ALS and the object co-ordinate systems are assumed to be identical. The stochastic model of an ALS point is described by two standard deviations: σ_{XY} for its planimetric co-ordinates and σ_Z for its height. The edge point co-ordinates are introduced with a standard deviation σ_E.

8.3.4 Overall Adjustment

All observations are used in an overall adjustment process. The weights of the observations are determined from their a priori standard deviations. Correlations between the observations (e.g. between the x and y image co-ordinates of an image point) are not considered. Robust estimation is carried out by iteratively re-weighting the observations depending on their normalised residuals in the previous adjustment [Kager, 2000]. The reweighting scheme is only applied to the sensor observations and to the observations modelling geometric constraints, in order to eliminate gross observation errors and wrong hypotheses about geometric regularities. The surface parameters and the vertex co-ordinates determined in the adjustment are used to derive the final building model.

8.4 Evaluation

8.4.1 The Test Data

For our test, we selected 8 buildings of different size and complexity out of a larger test area in Fairfield (NSW). They were chosen to highlight the method's potential to handle buildings of both regular and irregular shapes. Both ALS and image data were available for that test site. The ALS data were captured using an Optech ALTM 3025 laser scanner with a nominal average point distance of 1.25 m. As our test buildings were at the edge of a swath, there was a relatively irregular point density, with point distances of about 0.5 m in flight direction and 1.5 - 2 m across flight direction. The aerial images were a stereo pair taken at a scale of 1:11000 (focal length f = 30 cm). They were scanned at a resolution of 15 μm, which corresponds to a ground sampling distance of 0.17 m.

8.4.2 Generating Reference Data

The aerial images were used to determine the reference data for the test. In a semi-automatic working environment, the roof polygons were digitised in the images and hypotheses about geometric regularities were introduced by the human operator. The adjustment model described in section 8.3 was used to determine the parameters of the reference buildings, taking into account the GESTALT observations, the image co-ordinates of the building vertices, and ALS points to improve the height accuracy of the reference models. The ALS points were necessary because of the weak configuration of the images. Figure 8.2 shows an upright projection of a reference building resembling a hip roof and the ALS points. Three variants are shown: the results of photogrammetric plotting with and without geometric constraints and the results achieved by combining photogrammetric plotting with geometric constraints and ALS data. For the variant without geometric constraints the RMS values of the height differences of the horizontal eaves is $\pm$0.25 m. In the constrained version, the eaves are horizontal, but the figure reveals that the heights of the eaves derived from the ALS data are about 50 cm lower. The ALS points were introduced as model co-ordinates; the shift was about 15 cm in X and Y and about 5 cm in Z. The precision of the building vertices was about $\pm$17 cm in X and Y, and about $\pm$5 cm in Z.

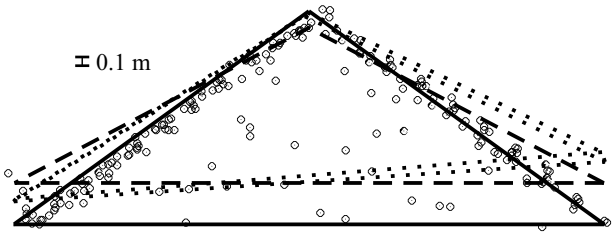


Figure 8.2: Upright projection of a hip roof (heights enlarged by a factor 2) generated from images without constraints (dotted lines); images with constraints (broken lines); images with constraints and ALS points (full lines). Circles: ALS points.

8.4.3 Results and Discussion

From the ALS data, a DSM with a grid width of Δ = 0.5 m was generated. From the DSM, roof planes were extracted, and the roof boundary polygons were determined as a combination of intersection lines and step edges in the way described in [Rottensteiner et al., 2005a]. These initial roof boundary polygons are shown super-imposed to the DSM in figure 8.3.

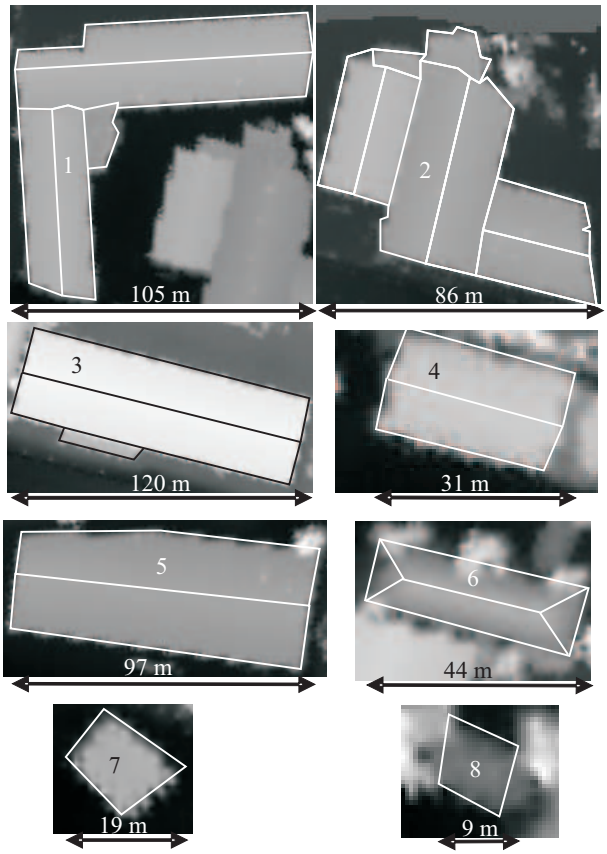


Figure 8.3: Initial roof boundary polygons for the eight buildings superimposed to the DSM. The buildings are shown in different scales, according to the extents shown in the figure.

In general the models look quite good except for building 8, which is partly occluded by trees. There is some noise in the outlines of buildings 1 and 2. Buildings 4, 6, 7, and 8 and the main part of building 3 should have a rectangular footprint, which is not entirely preserved in the initial models; geometric constraints should help to overcome this situation. The initial models, the original ALS points, and the step edge points provide the input for the overall adjustment. Soft constraints were introduced just on the basis of a comparison of angles/height differences to thresholds. Table 8.1 gives an overview about the stochastic model for the individual groups of observations in adjustment. Robust estimation was applied to the soft constraints and to the ALS and step edge points. In the current implementation this had to be done in a supervised way. It turned out that with some larger buildings the stochastic model had to be changed to make false hypotheses on geometric constraints detectable. Using σ_C = $\pm$0.05 m and σ_E = $\pm$0.25 m turned out to be a good choice. However, the final adjustment without the eliminated observations was carried out using the values given in table 8.1. They were confirmed by a variance component analysis.

Topology σ_T $[m]$	Constraints σ_C $[m]$	ALS XY σ_{XY} $[m]$	ALS Z σ_Z $[m]$	Step Edge σ_E $[m]$
$\pm$0.01	$\pm$0.015	$\pm$0.25	$\pm$0.075	$\pm$0.5

Table 8.1: A priori standard deviations of the observations.

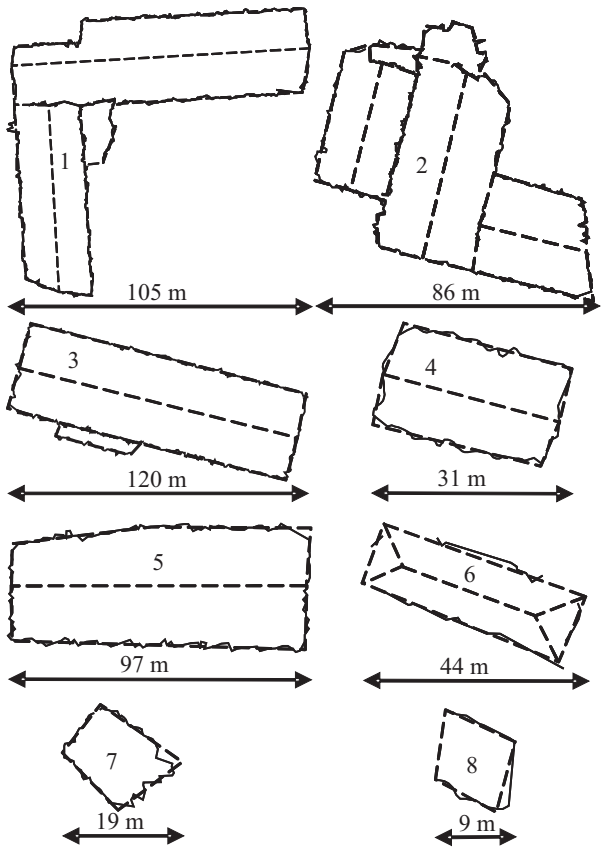


Figure 8.4: Final roof boundary polygons (red, broken lines) and reference data (blue). A part of building 2 is missing in the reference data since it only occurs in the ALS data. A colour version of this figure can be found on page 132.

Figure 8.4 gives the final results of building reconstruction and a comparison to the reference data. Compared to figure 8.3, the building models appear to be more regular. For buildings 1-6 the number of extracted roof planes was correct. The intersection lines are very accurate, and step edges are in general determined quite well, too. Some small roof structures are generalised, e.g. the outline of the smallest roof plane of building 1 or of roof plane a of building 2. The step edge between that plane and its neighbouring plane b was not determined very precisely, either. The problem was that roof plane a was horizontal, its western vertex being higher and its eastern vertex lower than the corresponding vertices of roof plane b; the maximum height difference was only 0.3 m, so that the step edge was poorly defined. Building 7 was reconstructed as being flat. The intersection of the two roof planes is only 0.15 m lower than the eaves, which is the reason why the two planes were merged. Building 8 was also reconstructed as a flat roof. It was the smallest building in the sample with only a few ALS points on the roof planes, and both ends occluded by trees. The outlines at the occluded

ends are not very well detected either. Apart from the visual inspection of the building models, a numerical evaluation of these results was carried out. RMS values of the co-ordinate differences of corresponding vertices in the reconstruction results and the reference data were computed for each roof plane:

$$RMS_{XY} = \sqrt{\frac{\sum(\Delta X^2 + \Delta Y^2)}{N}} \;\; and \;\; RMS_Z = \sqrt{\frac{\sum(\Delta Z^2)}{N}} \tag{8.5}$$

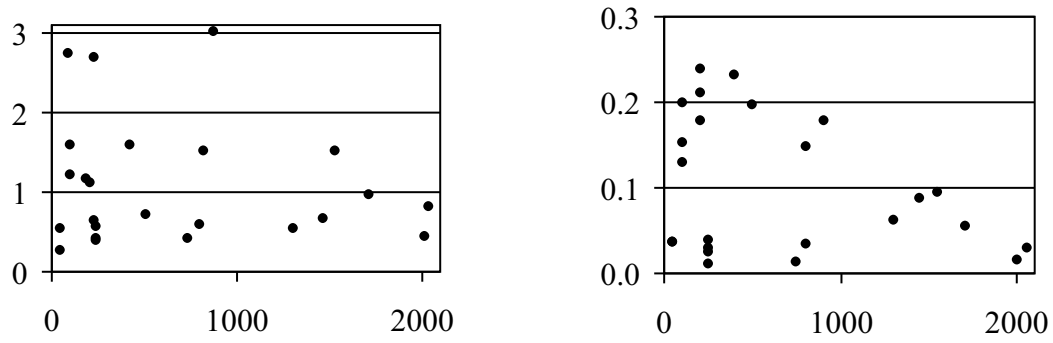


Figure 8.5: Left: RMS_{XY} $[m]$, right: RMS_Z $[m]$, both depending on the roof area $[m^2]$.

In equation 8.5, N is the number of corresponding points in the respective roof plane. If no matching vertex was found, the closest point on the corresponding roof boundary polygon was used instead. For buildings 7 and 8 only the outlines were evaluated. Figure 8.5 shows a graph of RMS_{XY} and RMS_Z depending on the roof area. RMS_{XY} is smaller than 3.1 m for all roof planes. For most roofs it is in the range between $\pm$0.5 m and $\pm$1.5 m, which is better than the point density across the flight direction. The largest values occur for roof planes smaller than 100 m^2, with the exception of roof plane a and b of building 2, for reasons discussed above. RMS_Z is much smaller than RMS_{XY} because heights are better defined in ALS data than step edges. RMS_Z becomes smaller with increasing area roof planes because more ALS points give support to large planes. Intersections are more accurately determined than step edges. RMS values computed for intersection lines are only $\pm$0.35 m in planimetry and $\pm$0.07 m in height.

B	P	RMS_{XY} $[m]$	RMS_Z $[m]$	Δ_{XY} $[m]$	Δ_Z $[m]$
1	5	0.76	0.12	0.24	0.01
2	5	2.27	0.20	0.00	-0.02
3	3	0.82	0.10	0.07	0.16
4	2	0.60	0.02	0.13	0.03
5	2	1.31	0.08	-0.08	-0.02
6	4	0.48	0.09	0.36	0.17
7	2	1.43	0.14	0.44	0.03
8	O	2.74	-	-0.02	-

Table 8.2: B: Building; P: Number of planes; RMS_{XY}, RMS_Z: Combined RMS values in planimetry / height; Δ_{XY}, Δ_Z: improvement of RMS_{XY} / RMS_Z.

Table 8.2 gives combined RMS values for all the test buildings. The large value for RMS_{XY} for building 2 of $\pm$2.27 m is caused by the erroneous step edge; the combined value without that edge would be $\pm$1.43 m. For most buildings, RMS_{XY} is better than the average point distance across flight direction. Apart from problems with low step edges, errors occurred at the outlines of some of the larger building due to occlusions: as the test area was at the edge of the swath, the positions of the step edges were very uncertain there. The height accuracy is good, with the largest value of $\pm$0.20 m occurring at building 2, again at the problematic step edge. Table 8.2 also gives the impact of the overall adjustment to the RMS values. With building 5, the RMS values get worse by a small value after adjustment, but in most cases the RMS values are improved by the overall adjustment. The improvement can be up to 45 % (building 6).

8.5 Conclusion

In this chapter we have described a model for the consistent estimation of building parameters that is part of a method for the automatic reconstruction of buildings from ALS data. The adjustment model can consider geometric regularities by "soft constraints", and it can handle different sensor data. It was used not only in the reconstruction process, but also for the generation of reference data for a test project. In the test project, the roof boundary polygons extracted from the ALS data were compared to the reference data. The accuracy was determined to be in the range of or better than the average point distance in planimetry, and about $\pm$0.1 m - $\pm$0.2 m in height. The improvement of the model co-ordinates caused by the geometric constraints can be up to 45%.

Acknowledgements

The data set used in this study was provided by AAMHatch (www.aamhatch.com.au).

Bibliography

[Alharty and Bethel, 2004] Alharty, A. and Bethel, J. (2004). Detailed building reconstruction from airborne laser data using a moving surface method. In *Proceedings of the XX^{th} ISPRS Congress*, volume XXXV - B3 of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 213–218, Istanbul, Turkey.

[Ameri, 2000a] Ameri, B. (2000a). *Automatic recognition and 3D reconstruction of buildings from digital imagery*. PhD thesis, Institute of Photogrammetry, University of Stuttgart. Deutsche Geodätische Kommission Volume C-526.

[Ameri, 2000b] Ameri, B. (2000b). Feature based model verification (FBMV): A new concept for hypothesis validation in building reconstruction. In *Proceedings of the XIX^{th} ISPRS Congress*, volume XXXIII-B3 of *International Archives of Photogrammetry and Remote Sensing*, pages 24–35, Amsterdam, The Netherlands.

[Baillard et al., 1999] Baillard, C., Schmid, C., Zisserman, A., and Fitzgibbon, A. (1999). Automatic line matching and 3D reconstruction of buildings from multiple views. In Ebner, H., Eckstein, W., Heipke, C., and Mayer, H., editors, *Proceedings ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*, volume XXXII/3-2W5 of *International Archives of Photogrammetry and Remote Sensing*, pages 69–80, Munich, Germany.

[Binaghi et al., 1997] Binaghi, E., Gallo, I., Madella, P., and Rampini, A. (1997). Approximate reasoning and multistrategy learning for multisource remote sensing data interpretation. In Chen, C., editor, *Information processing for remote sensing*, pages 397–429, Singapore. World scientific publishing.

[Brenner, 2000] Brenner, C. (2000). *Erzeugung Gebäuderekonstruktion aus digitalen Oberflächenmodellen und Grundrissen*. PhD thesis, Institute of Photogrammetry, University of Stuttgart. Deutsche Geodätische Kommission Volume C-530.

[Brenner, 2005] Brenner, C. (2005). Constraints for modelling complex objects. In *Proceedings of the CMRT05 workshop*, volume XXXIV-3/W24 of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 49–54, Vienna, Austria.

[Briese, 2000] Briese, C. (2000). Digitale Modelle aus Laser-Scanner-Daten in städtischen Gebieten. Diploma Thesis, Institute of Photogrammetry and Remote Sensing, Vienna University of Technology.

[Briese et al., 2002] Briese, C., Pfeifer, N., and Dorninger, P. (2002). Applications of the robust interpolation for DTM determination. In *Proceedings of the ISPRS Commission III Symposium*, volume

XXXIV - 3A of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 55–61, Graz, Austria.

[Briese et al., 2001] Briese, C., Pfeifer, N., and Stadler, P. (2001). Derivation of digital terrain models in the SCOP++ environment. In *Proceedings of the OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for Detailed Digital Terrain Models*, Stockholm, Sweden.

[Brügelmann and Förstner, 1992] Brügelmann, R. and Förstner, W. (1992). Noise estimation for color edge extraction. In Förstner, W. and Ruwiedel, S., editors, *Robust Computer Vision*, pages 90–107, Karlsruhe, Germany. Wichmann.

[Brunn and Weidner, 1997] Brunn, A. and Weidner, U. (1997). Extracting buildings from digital surface models. In Baltavias, E., Eckstein, W., Gülch, E., Hahn, M., Stallmann, D., Tempfli, K., and Welch, R., editors, *Proceedings ISPRS Commission III/IV Workshop on 3D Reconstruction and Modelling of Topographic Objects*, volume XXXII/3-4W2 of *International Archives of Photogrammetry and Remote Sensing*, pages 27–34, Stuttgart, Germany.

[Förstner, 1994] Förstner, W. (1994). A framework for low level feature extraction. In Eklundh, J., editor, *Computer Vision - ECCV '94. Proceedings of the 5^{th} ICCV*, volume II, pages 383–394, Boston, MA.

[Fuchs, 1998] Fuchs, C. (1998). *Extraktion polymorpher Bildstrukturen und ihre topologische und geometrische Gruppierung*. PhD thesis, Institute of Photogrammetry, University of Bonn. Deutsche Geodätische Kommission Volume C-502.

[Geibel and Stilla, 2000] Geibel, R. and Stilla, U. (2000). Segmentation of laser altimeter data for building reconstruction: Different procedures and comparison. In *Proceedings of the XIX^{th} ISPRS Congress*, volume XXXIII-B3 of *International Archives of Photogrammetry and Remote Sensing*, pages 326–334, Amsterdam, The Netherlands.

[Gorte, 1999] Gorte, B. (1999). Supervised image classification. In Stein, A., van der Meer, F., and Gorte, B., editors, *Spatial statistics for remote sensing*, pages 153–163, Dordrecht, The Netherlands. Kluwer Academic Publishers.

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

[Haala et al., 1998] Haala, N., Brenner, C., and Anders, K.-H. (1998). Urban GIS from Laser Altimeter and 2D map data. In Schenk, T. and Habib, A., editors, *Proceedings ISPRS Commission III Symposium*, volume XXXII-3/1 of *International Archives of Photogrammetry and Remote Sensing*, pages 339–346, Columbus, OH.

[Heipke et al., 1997] Heipke, C., Mayer, H., Wiedemann, C., and Jamet, O. (1997). Evaluation of automatic road extraction. volume XXXII/3-2W3 of *International Archives of Photogrammetry and Remote Sensing*, pages 47–56.

[Hejmanowska, 1998] Hejmanowska, B. (1998). Removal of topographical effect from remote sensing data for thermal inertia modelling. In *Proceedings of the ISPRS Commission IV Symposium*, volume XXXII-4 of *International Archives of Photogrammetry and Remote Sensing*, pages 238–245.

[Heuel, 2004] Heuel, S. (2004). *Uncertain projective geometry. Statistical reasoning for polyhedral object reconstruction*. Springer-Verlag, Berlin Heidelberg, Germany, first edition.

[Hoover et al., 1996] Hoover, A., Gillian, J.-P., Jiang, X., Flynn, P. J., Bunke, H., Goldgof, D. B., Bowyer, K., Eggert, D. W., Fitzgibbon, A., and Fisher, R. B. (1996). An experimental comparison of range image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):673–689.

[Jiang and Bunke, 1999] Jiang, X. and Bunke, H. (1999). Edge detection in range images based on scan line approximation. *Computer Vision and Image Understanding*, 73(2):183–199.

[Kager, 1989] Kager, H. (1989). ORIENT: A universal photogrammetric adjustment system. In Grün, A. and Kahmen, H., editors, *Optical 3-D Measurement*, pages 447–455, Karlsruhe, Germany. Herbert Wichmann Verlag.

[Kager, 2000] Kager, H. (2000). Adjustment of algebraic surfaces by least squared distances. In *Proceedings of the XIX^{th} ISPRS Congress*, volume XXXIII-B3 of *International Archives of Photogrammetry and Remote Sensing*, pages 472–479, Amsterdam, The Netherlands.

[Klein, 1999] Klein, L. (1999). *Sensor and data fusion, concepts and applications*. SPIE Optical Engineering Press.

[Kraus, 2000] Kraus, K. (2000). *Photogrammetrie Band 3. Topographische Informationssysteme*. Dümmler Verlag, Köln, Germany, first edition.

[Kraus, 2002] Kraus, K. (2002). Principles of airborne laser scanning. *Journal of the Swedish Society of Photogrammetry and Remote Sensing*, 1:53–56.

[Kraus and Pfeifer, 1998] Kraus, K. and Pfeifer, N. (1998). Determination of terrain models in wooded areas with aerial laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53(4):193–203.

[Le Hégarat-Mascle et al., 1997] Le Hégarat-Mascle, S., Bloch, I., and Vidal-Madjar, D. (1997). Application of dempster-shafer evidence theory to unsupervised classification in multisource remote sensing. *IEEE Transactions on Geosciences and Remote Sensing*, 35(4):1018–1031.

[Lee et al., 1987] Lee, T., Richards, J. A., and Swain, P. (1987). Probabilistic and evidential approaches for multisource data analysis. *IEEE Transactions on Geosciences and Remote Sensing*, 25(3):283–293.

[Lu and Trinder, 2003] Lu, Y. and Trinder, J. (2003). Data fusion applied to automatic building extraction in 3D reconstruction. In *Proceedings of the Annual ASPRS conference*, pages 114–122, Anchorage, Alaska.

[Lu et al., 2006] Lu, Y., Trinder, J., and Kubik, K. (2006). Automatic building detection using the Dempster-Shafer algorithm. *Photogrammetric Engineering and Remote Sensing*, 72 (4):395–404.

[Maas, 1999] Maas, H.-G. (1999). Fast determination of parametric house models from dense airborne laserscanner data. In *International Workshop on Mobile Mapping Technology*, volume XXXII - 2W1 of *International Archives of Photogrammetry and Remote Sensing*, pages 1–6, Bangkok, Thailand.

[Matikainen et al., 2003] Matikainen, L., Hyyppä, J., and Hyyppä, H. (2003). Automatic detection of buildings from laser scanner data for map updating. In *ISPRS Workshop 3-D reconstruction from airborne laserscanner and InSAR data*, volume XXXIV - 3W13 of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 218–224, Dresden, Germany.

[McGlone, 1996] McGlone, C. (1996). Bundle adjustment with geometric constraints for hypothesis evaluation. In *Proceedings of the $XVIII^{th}$ ISPRS Congress*, volume XXXI-B3 of *International Archives of Photogrammetry and Remote Sensing*, pages 529–534, Vienna, Austria.

[Moons et al., 1998] Moons, T., Frere, D., Vandekerckhove, J., and Van Gool, L. (1998). Automatic modeling and 3D reconstruction of urban house roofs from high resolution aerial imagery. In *ECCV '98. Lecture notes in computer vision*, volume 1406(I), pages 410–425.

[Peternell and Steiner, 2004] Peternell, M. and Steiner, T. (2004). Reconstruction of piecewise planar objects from point clouds. *Computer-Aided Design*, 36:333–342.

[Pottmann et al., 2002] Pottmann, H., Leopoldseder, S., Wallner, J., and Peternell, M. (2002). Recognition and reconstruction of special surfaces from point clouds. In *Proceedings of the ISPRS Commission III Symposium*, volume XXXIV - 3A of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 271–276, Graz, Austria.

[Ragia and Winter, 2000] Ragia, L. and Winter, S. (2000). Contributions to a quality description of areal objects in spatial data sets. *ISPRS Journal of Photogrammetry and Remote Sensing*, 55(3):201–213.

[Rottensteiner, 2001] Rottensteiner, F. (2001). *Semi-automatic extraction of buildings based on hybrid adjustment using 3D surface models and management of building data in a TIS*. PhD thesis, Institute of Photogrammetry and Remote Sensing, Vienna University of Technology. Geowissenschaftliche Mitteilungen der TU Wien Nr 56.

[Rottensteiner, 2003] Rottensteiner, F. (2003). Automatic generation of high-quality building models from LIDAR data. *IEEE Computer Graphics and Applications*, 23(6):42–51.

[Rottensteiner, 2006] Rottensteiner, F. (2006). Consistent estimation of building parameters considering geometric regularities by soft constraints. In *Proceedings of the ISPRS Commission III Symposium*, volume XXXIV - 3 of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 13–18, Bonn, Germany.

[Rottensteiner and Briese, 2002] Rottensteiner, F. and Briese, C. (2002). A new method for building extraction in urban areas from high-resolution lidar data. In *Proceedings of the ISPRS Commission III Symposium*, volume XXXIV - 3A of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 295–301, Graz, Austria.

[Rottensteiner and Schulze, 2003] Rottensteiner, F. and Schulze, M. (2003). Performance evaluation of a system for semi-automatic building extraction using adaptable primitives. In *Proceedings of the ISPRS Conference on Photogrammetric Image Analysis (PIA)*, volume XXXIV/3-W8 of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 47–52, Munich, Germany.

[Rottensteiner et al., 2003] Rottensteiner, F., Trinder, J., Clode, S., and Kubik, K. (2003). Building detection using LIDAR data and multispectral images. In *Proceedings of the APRS Conference on Digital Image Computing: Techniques and Applications (DICTA)*, volume II, pages 673–682, Sydney, Australia.

[Rottensteiner et al., 2004] Rottensteiner, F., Trinder, J., Clode, S., and Kubik, K. (2004). Fusing airborne laser scanner data and aerial imagery for the automatic extraction of buildings in densely

built-up areas. In *Proceedings of the XX^{th} ISPRS Congress*, volume XXXV - B3 of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 512–517, Istanbul, Turkey.

[Rottensteiner et al., 2005a] Rottensteiner, F., Trinder, J., Clode, S., and Kubik, K. (2005a). Automated delineation of roof planes in LIDAR data. In *Laser Scanning 2005*, volume XXXIV-3/W19 of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 221–226, Enschede, The Netherlands.

[Rottensteiner et al., 2005b] Rottensteiner, F., Trinder, J., Clode, S., and Kubik, K. (2005b). Using the Dempster-Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion*, 6(4):283–300.

[Rottensteiner et al., 2007] Rottensteiner, F., Trinder, J., Clode, S., and Kubik, K. (2007). Building detection by fusion of airborne laserscanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(2):135–149.

[Schenk and Csatho, 2002] Schenk, T. and Csatho, B. (2002). Fusion of lidar data and aerial imagery for a more complete surface description. In *Proceedings of the ISPRS Commission III Symposium*, volume XXXIV - 3A of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 310–317, Graz, Austria.

[Stassopoulou et al., 2000] Stassopoulou, A., Caelli, T., and Ramirez, R. (2000). Automatic extraction of building statistics from digital Orthophotos. *International Journal of Geographical Information Science*, 14 (8):795–814.

[Steinle and Vögtle, 2000] Steinle, E. and Vögtle, T. (2000). Effects of different laser scanning modes on the results of building recognition and reconstruction. In *Proceedings of the XIX^{th} ISPRS Congress*, volume XXXIII-B3 of *International Archives of Photogrammetry and Remote Sensing*, pages 858–865, Amsterdam, The Netherlands.

[Sze et al., 1998] Sze, C.-J., Mark Liao, H.-Y., Hung, H.-L., Fan, K.-C., and Hsieh, J.-W. (1998). Multi-scale edge detection on range images via normal changes. *IEEE TCAS-II*, 45(8):1087–1092.

[Vallet and Taillandier, 2005] Vallet, B. and Taillandier, F. (2005). Fitting constrained 3D models in multiple aerial images. In *Proceedings of the British Machine Vision Conference*. Accessed in the WWW 19 07 2006: http://www.bmva.ac.uk/bmvc/2005/papers/paper-57-176.html.

[Vögtle and Steinle, 2000] Vögtle, T. and Steinle, E. (2000). 3D modelling of buildings using laser scanning and spectral information. In *Proceedings of the XIX^{th} ISPRS Congress*, volume XXXIII-B3 of *International Archives of Photogrammetry and Remote Sensing*, pages 927–933, Amsterdam, The Netherlands.

[Vögtle and Steinle, 2003] Vögtle, T. and Steinle, E. (2003). On the quality of object classification and automated building modeling based on laserscanning data. In *ISPRS Workshop 3-D reconstruction from airborne laserscanner and InSAR data*, volume XXXIV - 3W13 of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 149–155, Dresden, Germany.

[Vosselman, 1999] Vosselman, G. (1999). Building reconstruction using planar faces in very high density height data. In Ebner, H., Eckstein, W., Heipke, C., and Mayer, H., editors, *Proceedings ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*, volume XXXII/3-2W5 of *International Archives of Photogrammetry and Remote Sensing*, pages 87–92, Munich, Germany.

[Vosselman, 2002] Vosselman, G. (2002). On the estimation of planimetric offsets in laser altimetry data. In *Proceedings of the ISPRS Commission III Symposium*, volume XXXIV - 3A of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 375–380, Graz, Austria.

[Vosselman and Dijkman, 2001] Vosselman, G. and Dijkman, S. (2001). 3D building model reconstruction from point clouds and ground plans. In *Land surface mapping and characterization using laser altimetry*, volume XXXIV - 3W4 of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 37–43, Annapolis, MD, USA.

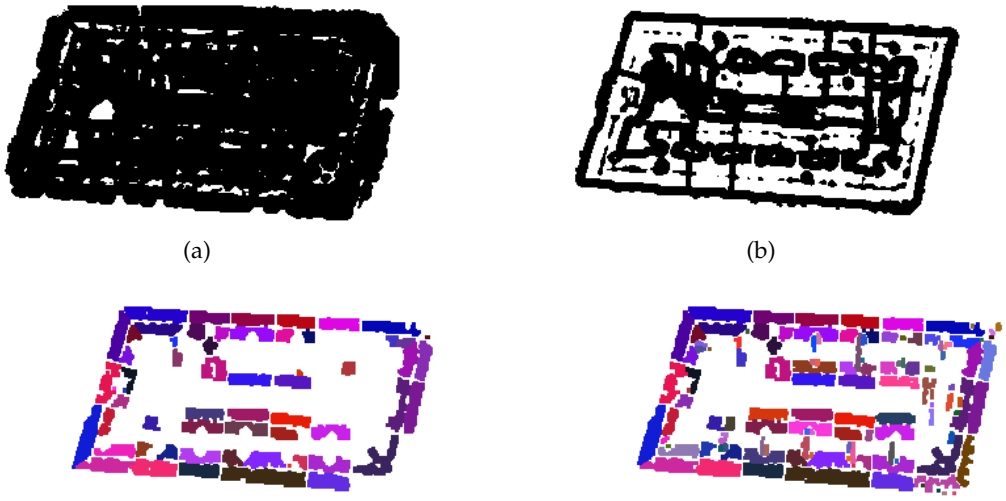
[Vosselman et al., 2004] Vosselman, G., Gorte, B., and G., S. (2004). Change detection for updating medium scale maps using laser altimetry. In *Proceedings of the XX^{th} ISPRS Congress*, volume XXXV - B3 of *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 207–212, Istanbul, Turkey.

[Weidner, 1997] Weidner, U. (1997). *Gebäudeerfassung aus digitalen Oberflächenmodellen*. PhD thesis, Institute of Photogrammetry, University of Bonn. Deutsche Geodätische Kommission Volume C-474.

[Weidner and Förstner, 1995] Weidner, U. and Förstner, W. (1995). Towards automatic building reconstruction from high resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4):38–49.

Appendix: Colour Figures

Figure 3.7: Boundary polygons super-imposed to the orthophoto.



(a) (b)

(c) (d)

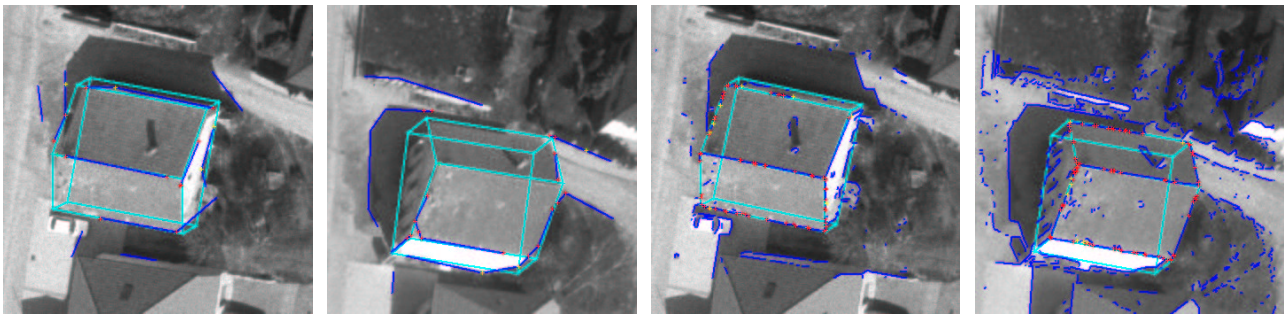
Figure 6.2: Detection of planar segments from a digital surface model in one of the building regions from figure 6.1. (a) and (b) show classification results of polymorphic feature extraction using different significance levels (first and last iteration). (c) Planar regions obtained by applying a connected component algorithm and region growing to (a). (d) Segment label image after the last iteration.

(a) (b) (c)

Figure 6.3: Postprocessing of the segmentation results: (a) unclassified pixels in the building region; (b) segment label image after post-processing containing additional planar segments; and (c) unclassified pixels after postprocessing.

(a) (b)

Figure 6.4: (a) Voronoi diagram of the label image in figure 6.3(b) and (b) Voronoi diagram after merging coplanar segments.



(a) (b) (c) (d)

Figure 6.11: Wire frame fitting of a building primitive resembling a saddleback roof in two images at two resolution levels of the image pyramids. Blue is the extracted image edges, red is the edge segment end points accepted as matches, yellow is the false matches, and cyan is the final positions of the wire frame.

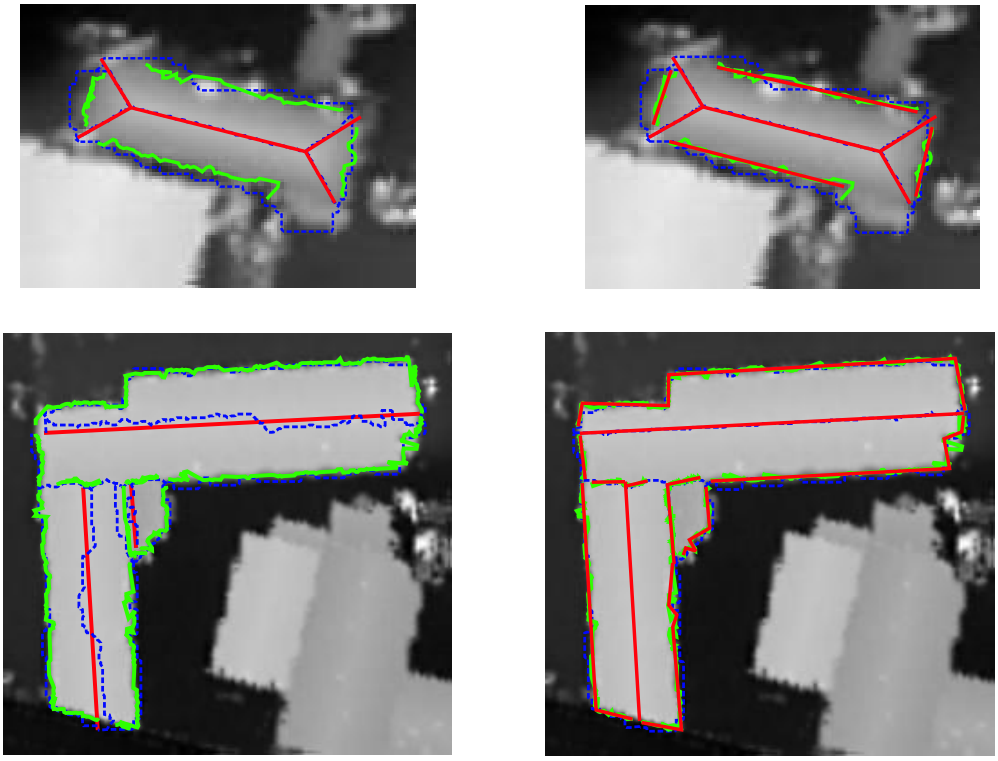


Figure 7.3: Blue dashed lines: approximate boundary polygons. Green: original step edges. Red: final polygon segments. Left: before generalisation of step edges. Right: after improving the planar segmentation.

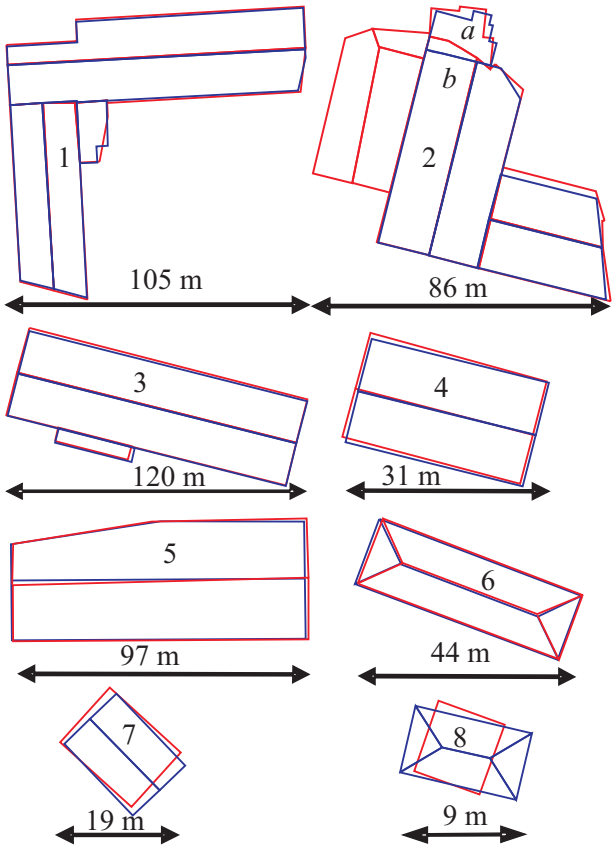


Figure 8.4: Final roof boundary polygons (red, broken lines) and reference data (blue). A part of building 2 is missing in the reference data since it only occurs in the ALS data.

GEOWISSENSCHAFTLICHE MITTEILUNGEN

Bisher erschienen:

Heft 1	Kolloquium der Assistenten der Studienrichtung Vermessungswesen. 1970-1973, Dezember 1976.
Heft 2	EGGER-PERDICH-PLACH-WAGENSOMMERER, Taschenrechner HP 45 und HP 65, Programme und Anwendungen im Vermessungswesen. 1. Auflage, März 1974, Special Edition in English Juli 1974, 2. verbesserte Auflage, November 1974.
Heft 3	Kolloquium der Assistenten der Studienrichtung Vermessungswesen. 1973-1974, September 1974.
Heft 4	EGGER-PALFINGER-PERDICH-PLACH-WAGENSOMMERER, Tektronix-Tischrechner TEK 31, Programmbibliothek für den Einsatz im Vermessungswesen, November 1974.
Heft 5	K. LEDERSTEGER, Die horizontale Isostasie und das isostatische Geoid, Februar 1975.
Heft 6	F. REINHART, Katalog von FK4 Horrebow-Paaren für Breiten von +30 bis +60, Oktober 1975.
Heft 7	Arbeiten aus dem Institut für Höhere Geodäsie, Wien, Dezember 1975.
Heft 8	Veröffentlichungen des Institus für Photogrammetrie zum XIII. Internationalen Kongreß für Photogrammetrie in Helsinki 1976, Wien, Juli 1976.
Heft 9	W. PILLEWIZER, Felsdarstellung aus Orthophotos, Wien, Juni 1976.
Heft 10	PERDICH-PLACH-WAGENSOMMERER, Der Einsatz des programmierbaren Taschenrechners Texas Instruments SR-52 mit Drucker PC100 in der ingenieurgeodätischen Rechentechnik, Wien, Mai 1976.
Heft 11	Kolloquium der Assistenten der Studienrichtung Vermessungswesen. 1974-1976, November 1976.
Heft 12	Kartographische Vorträge der Geodätischen Informationstage 1976, Wien, Mai 1976.
Heft 13	Veröffentlichungen des Institus für Photogrammetrie anläßlich des 80. Geburtstages von Prof. Dr. h.c. K. Neumaier, Wien, Januar 1978.
Heft 14	L. MOLNAR, Self Checking Analytical Relative Orientation and Strip Formation, Wien, Dezember 1978.
Heft 15	Veröffentlichungen des Institus für Landesvermessung anläßlich des 80. Geburtstages von Prof. Dr. Alois Bavir, Wien, Januar 1979.
Heft 16	Kolloquium der Assistenten der Studienrichtung Vermessungswesen. 1976-1978, November 1979.
Heft 17	E. VOZIKIS, Die photographische Differentialumbildung gekrümmter Flächen mit Beispielen aus der Architekturbildmessung, Wien, Dezember 1979.
Heft 18	Veröffentlichungen des Institus für allgemeine Geodäsie anläßlich des 75. Geburtstages von Prof. Dipl.-Ing. Dr. F. Hauer, Die Höhe des Großglockners, Wien, 1981.
Heft 19	H. KAGER, Bündeltriangulation mit indirekt beobachteten Kreiszentren, Wien, April 1981.
Heft 20	Kartographische Vorträge der Geodätischen Informationstage 1980, Wien, Mai 1982.
Heft 21	Veröffentlichungen des Institus für Kartographie anläßlich des 70. Geburtstages von Prof. Dr. Wolfgang Pillewizer: Glaziologie und Kartographie, Wien, Dezember 1982.
Heft 22	K. TEMPFLI, Genauigkeitsschätzung digitaler Höhenmodelle mittels Spektralanalyse, Wien, Mai 1982.
Heft 23	E. CSAPLOVICS, Interpretation von Farbinfrarotbildern, Wien, November 1982.

Heft 24 J. JANSA, Rektifizierung von Multispektral-Scanneraufnahmen - Entwicklung und Erprobung eines EDV-Programms, Wien, Mai 1983.

Heft 25 Zusammenfassung der Diplomarbeiten, Dissertationen und Habilitationen an den geodätischen Instituten der TU Wien, Wien, November 1984.

Heft 26 T. WUNDERLICH, Die voraussetzungsfreie Bestimmung von Refraktionswinkeln, Wien, August 1985.

Heft 27 G. GERSTBACH (Hrsg.), Geowissenschaftliche/geotechnische Daten in Landinformationssystemen - Bedarf und Möglichkeiten in Österreich, Juni 1986.

Heft 28 K. NOVAK, Orientierung von Amateuraufnahmen ohne Paßpunkte, Wien, August 1986.

Heft 29 Veröffentlichungen des Institus für Landesvermessung und Ingenieurgeodäsie, Abt. Ingenieurgeodäsie, anläßlich des 80. Geburtstages von Prof. Dipl.-Ing. Dr. F. Hauer, Wien, Oktober 1986.

Heft 30 K.-H. ROCH, Über die Bedeutung dynamisch ermittelter Parameter für die Bestimmung von Gesteins- und Gebirgseigenschaften, Wien, Februar 1987.

Heft 31 G. HE, Bildverbesserung mittels digitaler Filterung, Wien, April 1989.

Heft 32 F. SCHLÖGELHOFER, Qualitäts- und Wirtschaftlichkeitsmodelle für die Ingenieurphotogrammetrie, Wien, April 1989.

Heft 33 G. GERSTBACH (Hrsg.), Geowissenschaftliche/geotechnische Daten in Landinformationssystemen - Datenbestände und Datenaustausch in Österreich, Wien, Juni 1989.

Heft 34 F. HOCHSTÖGER, Ein Beitrag zur Anwendung und Visualisierung digitaler Geländemodelle, Wien, Dezember 1989.

Heft 35 R. WEBER, Lokale Schwerefeldmodellierung unter Berücksichtigung spektraler Methoden zur Geländereduktion, Wien, April 1990.

Heft 36 o. Prof. Dr. Hans Schmid zum 70. Geburtstag. Veröffentlichung der Abteilung für Landesvermessung, Wien, Oktober 1990.

Heft 37 G. GERSTBACH, H. P. HÖLLRIEGL und R. WEBER, Geowissenschaftliche Informationsbörse - Eine Nachlese zur GeoLIS II, Wien, Oktober 1990.

Heft 38 R. ECKER, Rastergraphische Visualisierungen mittels digitaler Geländemodelle, Wien, August 1991.

Heft 39 Kartographische Forschungen und anwendungsorientierte Entwicklungen, herausgegeben von W. Stams und F. Kelnhofer zum 80. Geburtstag von Prof. Dr. W. Pillewizer, Wien, Juli 1991.

Heft 39a W. RIEGER, Hydrologische Anwendungen des digitalen Geländemodelles, Wien, Juli 1992.

Heft 40 K. STEINNOCHER, Methodische Erweiterungen der Landnutzungsklassifikation und Implementierung auf einem Transputernetzwerk, Wien, Juli 1994.

Heft 41 G. FORKERT, Die Lösung photogrammetrischer Orientierungs- und Rekonstruktionsaufgaben mittels allgemeiner kurvenförmiger Elemente, Wien, Juli 1994.

Heft 42 M. SCHÖNER, W. SCHÖNER, Photogrammetrische und glaziologische Untersuchungen am Gsbre (Ergebnisse der Spitzbergenexpedition 1991), Wien, Mai 1996.

Heft 43 M. ROIC, Erfassung von nicht signalisierten 3D-Strukturen mit Videotheodoliten, Wien, April 1996.

Heft 44 G. RETSCHER, 3D-Gleiserfassung mit einem Multisensorsystem und linearen Filterverfahren, Wien, April 1996.

Heft 45 W. DAXINGER, Astrogravimetrische Geoidbestimmung für Ingenieurprojekte, Wien, Juli 1996.

Heft 46 M. PLONER, CCD-Astrometrie von Objekten des geostationären Ringes, Wien, November 1996.

Heft 47 Zum Gedenken an Karl Killian “Ingenieur" und “Geodät" 1903-1991, Veröffentlichung der Fachgruppe Geowissenschaften, Wien, Februar 1997.

Heft 48 A. SINDHUBER, Ergänzung und Fortführung eines digitalen Landschaftsmodelles mit multispektralen und hochauflösenden Fernerkundungsaufnahmen, Wien, Mai 1998.

Heft 49 W. WAGNER, Soil Moisture Retrieval from ERS Scatterometer Data, Wien, Dezember 1998.

Heft 50 R. WEBER, E. FRAGNER (Editoren), Prof. Bretterbauer, Festschrift zum 70. Geburtstag, Wien, Juli 1999.

Heft 51 Ch. ÖHRENEDER, A Similarity Measure for Global Image Matching Based on the Forward Modeling Principle, Wien, April 1999.

Heft 52 M. LECHTHALER, G. GARTNER (Hrsg.), Per Aspera ad Astra, Festschrift für Fritz Kelnhofer zum 60. Geburtstag, Wien, Jänner 2000.

Heft 53 F. KELNHOFER, M. LECHTHALER (Hrsg.), Interaktive Karten (Atlanten) und Multimedia-Applikationen, Wien, März 2000.

Heft 54 A. MISCHKE, Entwicklung eines Videotheodolit-Meßsystems zur automatischen Richtungsmessung von nicht signalisierten Objektpunkten, Wien, Dezember 2000.

Heft 55 Veröffentlichung des I.P.F. anlässlich der Emeritierung von Prof. Dr. Peter Waldhäusl, Wien, 2001.

Heft 56 F. ROTTENSTEINER, Semi-automatic Extraction of Buildings Based on Hybrid Adjustment Using 3D Surface Models and Management of Building Data in a TIS, Wien, Juni 2001.

Heft 57 D. LEGENSTEIN, Objektrekonstruktion aus perspektiven Bildern unter Einbeziehung von Umrisslinien, Wien, Mai 2001.

Heft 58 F. KELNHOFER, M. LECHTHALER und K. BRUNNER (Hrsg.), Telekartographie und Location Based Services, Wien, Jänner 2002.

Heft 59 K. BRETTERBAUER, Die runde Erde eben dargestellt: Abbildungslehre und sphärische Kartennetzentwürfe, Wien, 2002.

Heft 60 G. GARTNER, Maps and the Internet 2002, Wien 2002.

Heft 61 L. DORFFNER, Erzeugung von qualitativ hochwertigen 3D Photomodellen für Internetbasierte Anwendungen mit besonderem Augenmerk auf Objekte der Nahbereichsphotogrammetrie, Wien, Jänner 2002.

Heft 62 K. CHMELINA, Wissensbasierte Analyse von Verschiebungsdaten im Tunnelbau, Wien, 2002.

Heft 63 A. NIESSNER, Qualitative Deformationsanalyse unter Ausnutzung der Farbinformation, Wien, 2002.

Heft 64 K. BRETTERBAUER; R. WEBER, A Primer of Geodesy for GIS-Users, Wien, im Herbst 2003.

Heft 65 N. PFEIFER, 3D Terain Models on the basis of a triangulation, Wien, Jänner 2002.

Heft 66 G. GARTNER (Hrsg), Location Based Services & Telecartography, Wien, 2004.

Heft 67 I. KABASHI, Gleichzeitig-gegenseitige Zenitwinkelmessung über größere Entfernungen mit automatischen Zielsystemen, Wien, 2004.

Heft 68 J. BÖHM, Troposphärische Laufzeitverzögerungen in der VLBI, Wien, 2004.

Heft 69 R. WEBER, W. SCHLÜTER, U. SCHREIBER, O. TITOV, Evolving Space Geodesy Techniques (EGS XXVII General Assembly, Nice, France, 2002), Wien, 2004.

Heft 70 G. WEINWURM, Amalthea's Gravity Field and its Impact on a Spacecraft Trajectory, Wien 2004.

Heft 71 Forschungsgruppe Ingenieurgeodäsie, Festschrift anlässlich des 65. Geburtstages von Herrn o.Univ.Prof.Dr.-Ing. Heribert Kahmen, Wien, 2005.

Heft 72 A. REITERER, A Knowledge-Based Decision System for an On-Line Video-Theodolite-Based Multisensor System, Wien, 2005.

Heft 73 M. HABERLER, Einsatz von Fuzzy Methoden zur Detektion konsistenter Punktbewegungen, Wien, 2005.

Heft 74 G. GARTNER, Location Based Services & Telecartography, Proceedings of the Symposium 2005, Wien, 2005.

Heft 75 Th. HOBIGER, VLBI as a tool to probe the ionosphere, Wien, 2006.

Heft 76 E. KLAFFENBÖCK, Troposphärische Laufzeitverzögerung von GNSS-Signalen - Nutzen aktiver Referenzstationsnetze für die Meteorologie, Wien, 2006.

Heft 76a P.J. MENDES-CERVEIRA, Tidal and non-tidal contributions to surface loading processes on station coordinates, Wien, 2006.

Heft 78 G. KOSTOV, G. BOURDA, L. FERNANDEZ, T. KONDO, Research Projects at IGG Reports, Wien, 2007.

Heft 79 J. BÖHM, A. PANY, H. SCHUH (Editors), Proceedings of the 18th European VLBI for Geodesy and Astrometry Working Meeting, 12-13 April 2007, Wien, 2007.

Heft 80 J. BÖHM, Tropospheric Delay Modelling at Radio Wavelengths for Space Geodetic Techniques, Wien, 2007.

Heft 81 G. Retscher, Mobile Multi-sensor Systems for Personal Navigation and Location-based Services, Wien, 2007.

Heft 82 R. HEINKELMANN, Bestimmung des atmosphärischen Wasserdampfes mittels VLBI als Beitrag zur Klimaforschung, Wien, 2008.