Doctoral Dissertation

CHARACTERIZATION OF REAL PROPERTIES USING 3D VISION

PHILIPP MEIXNER

Graz University of Technology/ Technische Universität Graz

Faculty of Computer Science/ Fakultät für Informatik

Institute for Computer Graphics and Vision/ Institut für Maschinelles Sehen und Darstellen

Referees

Prof. Dr. Franz Leberl

Thesis Supervisor

Graz, University of Technology

Faculty of Computer Science

Prof. Dr. Nicolas Paparoditis

Referee

Institut Géographique National, Université Paris-Est

Laboratoire MATIS

Graz, Austria, May 2012

"All truths are easy to understand once they are discovered; the point is to discover them."

Galileo Galilei

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

……..	……..	……..
Ort	Datum	Unterschrift

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

……..	……..	……..
Place	Date	Signature

ACKNOWLEDGEMENT

Writing a doctoral thesis is a not an easy endeavor as learned during the past 3 years. All who have embarked on such a journey understand the challenges. Although I was responsible for completing my thesis, I would not have been able to do so successfully without the support, collaboration, comments and critiques from others.

First of all, I would like to express my deepest gratitude to my supervisor Prof. Franz Leberl from the Institute for Computer Graphics and Vision who gave me the opportunity to pursue a PhD at his prestigious Institute. Since the beginning, Prof. Franz Leberl has always encouraged my scientific research in the exciting field of Computer Vision and provided me with many innovative and inspiring ideas. Without his great support, I would not have been able to do my research and to write this thesis. Special thanks also go to Prof. Horst Bischof, who supported me, when Prof. Franz Leberl was engaged with other projects.

While working on my thesis project, I was employed as a researcher with fundings from the VRVIS Research Center for Virtual Reality and Visualization (VRVIS) in Vienna. I want to thank VRVIS for the collaboration, support and for various opportunities such as conference participations.

I especially would like to thank my second supervisor, Nicolas Paparoditis from the laboratoire MATIS at the Institute Géographique Nationale (IGN), who invited me for several months to his exciting Institute where "urban 3D models" have been investigated since 1992. I also want to thank all my former colleagues at MATIS, with a special thanks to Dr. Mathieu Brédif, Dr. Bahman Soheilian and Dr. Clement Mallet for their great support and collaboration. Through them I gained appreciation for the difficulties of modeling building roofs.

I would like to express my appreciation to all the doctoral students, doctoral graduates and postdocs from the ICG. Special thanks go to Stefan Kluckner, Arnold Irschara, Thomas Pock, Michael Donoser, Hayko Riemenschneider, Matthias Rüther and Markus Unger for their help and patience in answering my many questions. Their help, comments and critiques were a big inspiration for my work. Coming from a background of Surveying Engineering, I depended very much on their help to cope with a focus on computer science, software systems, algorithms and computer vision. I also want to thank everybody for sharing ideas and corporation to generate papers with me.

I would like to thank the people working at Microsoft Photogrammetry, Dr. Michael Gruber and Dr. Barbara Gruber-Geymayer, for their support and the inspiring discussions that led to further research directions. They provided data, insights into image classifications, accuracies and coordinate systems

Above all, I want to thank my father and my mother for providing me with the means to acquire my doctorate, which included their motivation, encouragement and financial support.

ABSTRACT

Accurate and realistic 3-dimensional models of urban environments are increasingly important for applications like virtual tourism, city planning, Internet search and many emerging opportunities in the context of "ambient intelligence". Applications like Bing-Maps or Google Earth are offering virtual models of many major urban areas worldwide. While initially these data sets support visualization, they are inherently capable of addressing a broader purpose. On the horizon are urban models that consist of semantically interpreted objects; an urban 3D visualization will be computer generated, with a fundamental advantage: the urban models can be searched based on object classes. This characterization is the perquisite for a property valuation that we want to enable by semantically interpreting these models.

Oblique aerial photography and LiDAR have become a widely used resource for urban imaging and modeling purposes. Originating in the US and championed by Pictometry, oblique images are now being acquired world-wide. We are interested in a comparison between oblique and vertical aerial photography, especially addressing the facades in urban areas and facade details such as the number of floors and windows. Our results show that vertical imagery is well-suited to facade analysis, and that oblique images deliver results compromised by occlusions. This indicates that the benefit of oblique images is questionable in cases where high overlapping vertical images exist.

We are also interested in a comparison between LiDAR data and vertical aerial images, especially addressing the roofs in urban areas and their characterization. We show that the photogrammetric accuracy compares well with LiDAR, yet the density of surface points is much higher from images.

This project presents a framework which specifies the processing steps that are necessary for a reasonable semantic interpretation. Our focus is on characterizing the individual buildings on a real property. A key to success in this task is the availability of aerial photography at a greater overlap than has been customary in traditional photogrammetry, as well as a Ground Sampling Distance (GSD) exceeding the traditional values.

We first describe the different source data which have to be brought into a common coordinate system. In this process, we build an integrated geometry and semantic object data set that can be analyzed for various purposes.

We start out by merging the aerial imagery with the cadastral information to define each property as a separate entity for further analysis. The cadastral data may also contain preliminary information about a building footprint. In cases of missing cadastral data, we also present a novel method for facade separation from aerial images with a success rate of 88%.

An important aspect on each real property are its buildings. Therefore, in the next step the building footprints get refined vis-à-vis the mere cadastral prediction, using an image classification and the definition of roof lines. 3-D facade coordinates are computed from aerial image segments, the cadastral information, and the DTM.

We describe how one can recognize and reconstruct buildings and their facades in 3 dimensions with the purpose of extracting the building size, its footprint, the number of floors, the roof shapes, the number and size of windows, and the existence or absence of balconies. Since facades are often not simply vertical planes, but have depth in the form of stairwells, balconies, awnings or decorations, we need to consider a 3D approach. Mapping a vertical object such as a facade by means of vertically looking aerial photography requires one to cope with very steep look angles and pixels far from square in object space. We can conclude that current high overlaps in imaging campaigns support the creation of useful 3D point clouds of vertical facades. We achieved accuracies sufficient to count windows and floors, even when the buildings had complex shapes with stairwells or balconies. In a test area in Graz (Austria) with 216 buildings, we found that planar facades could be interpreted with success rates between 86% and 93%. When dealing with complex facades the problem gets more difficult. Without the use of the 3rd dimension none of these facades could be interpreted correctly. By including this information, 86% of all floors and 80% of all windows could be counted correctly.

However, our focus not only lies in describing the facades of a building but also its roof. Building roofs have, in the last couple of years, been studied with aerial LiDAR point clouds (Light Detection And Ranging). However, recent progress in digital aerial cameras has rendered possible the acquisition of very dense point clouds from high overlap digital aerial imagery, and to use these point clouds jointly with the image information to generate 3D building models.

This project presents a multi-step processing framework and work flow for the automatic segmentation and 3D reconstruction of building roofs in densely built-up areas from high-resolution vertical aerial images. We are interested in two major topics: the determination of the roof shape and the detection of superstructures like dormer windows, chimneys and skylight to determine the degree of extension of a rooftop.

For the determination of the roof shape details extruding from, or intruding into, a roof is being excluded so that each roof is being modeled by means of its planar segments and can then be classified as a specific roof type from a set of standard roof shapes.

We show that the results from aerial photography compete well with LiDAR-results as reported by LiDAR researchers. We evaluated three different plane detection techniques and compared their results. In real city data with a total of 1069 superstructures the proposed automated process finds 1015, thus achieving success at a rate of 95%. For our Graz and Annecy dataset with 216 respectively 35 buildings with 810 and 159 roof planes, results in correct roof planes in 95% and 91% of all cases. Small roof structures do confuse the analysis and must therefore be detected and eliminated. These structures can be detected with success rates of 95% respectively 86% of all the test cases depending on the accuracy of the range data. Roof types get classified correctly at a rate of 85%. We also show that we can reconstruct the building roofs and their superstructures such as chimneys and dormer windows with high accuracy and completeness because of the high point density of 100 pixels per m² in contrast to 25 points per m² using conventional LIDAR sensors. Concerning the determination of the degree of extension of a roof we achieve a detection rate of 87%.

We demonstrate in this project that vertical aerial images are a useful and economical alternative to LiDAR and oblique aerial images for the characterization of building facades and

roofs. Our experiments are evaluated on a variety of datasets (Graz, Annecy). Our experiments demonstrate robustness and high geometric accuracy of the proposed methods and show that it is feasible to use this information for valuation purposes.

KURZFASSUNG

Genaue und realistische 3D Modelle von städtischen Bereichen gewinnen immer mehr Bedeutung im Bereich der Stadtplanung, des Tourismus, der Internetsuche und im Bereich der „Ambient Intelligence“. Anwendungen wie Bing Maps oder Google Earth bieten diese 3D Modelle für viele Städte weltweit an. Lag der Hauptanwendungszweck ursprünglich im Bereich der Visualisierung, eröffnen sich durch den technischen Fortschritt neue Anwendungsbereiche, wie z.B. die semantische Interpretation dieser Modelle. Computer generieren dabei diese Modelle vollautomatisch und werten diese je nach Verwendungszweck aus. Das Hauptziel dieser Arbeit ist es diese Modelle so zu charakterisieren, dass sie für eine Liegenschaftsbewertung verwendet werden können.

Schrägaufnahmen und LiDAR sind seit Jahren für die Generierung von 3D Stadtmodellen und deren Interpretation die Hauptdatenquellen. Wir führen einen Vergleich zwischen diesen Datenquellen und vertikalen Luftbildern durch und zeigen, dass besonders im Bereich der 3D Stadtmodellierung und Interpretation vertikale Luftbilder große Vorteile besitzen. Der technische Fortschritt im Bereich der Luftbildfotographie und die dadurch entstehenden hochauflösenden und hoch überlappenden Bilder bilden die Grundlage dafür.

Der Hauptfokus der Arbeit liegt in der semantischen Interpretation von Gebäuden. Wir beschreiben, wie Gebäude, deren Fassaden und Dächer detektiert und 3D rekonstruiert werden können. Dabei sind Merkmale wie die Anzahl der Stockwerke, die Anzahl und Größe der Fenster, die Detektion von Balkonen und Vorsprüngen, die Erkennung von Rauchfängen, Dachgauben, die Bestimmung der Dachformen und des Ausbaugrades des Daches von Interesse. Da wir es nicht nur mit planaren Fassaden, sondern oft mit komplexen Fassaden mit Stiegenhäusern, Balkonen oder Erkern zu tun haben, präsentieren wir eine Methode zur 3 dimensionalen Rekonstruktion von Fassaden, ohne die eine sinnvolle Interpretation nicht möglich wäre.

Es ist möglich im Falle von planaren Fassaden Stockwerke und Fenster mit einer Genauigkeit von 86% bzw. 95% zu detektieren. Im Falle von komplexen Fassaden erhalten wir eine Genauigkeit von 80% bzw. 86%, die ohne einer 3D Rekonstruktion nicht interpretierbar wären.

Allerdings liegt unser Interesse nicht nur bei Fassaden sondern auch bei Dächern. Wir präsentieren ein Multi-Step Processing Framework zur automatischen Klassifizierung und Rekonstruktion von Dächern. Dabei sind die Dachform, der Dachausbau, Rauchfänge sowie Dachgauben von Interesse. Im Rahmen der Arbeit haben wir 3 verschiedene Verfahren zur Ebenendetektion anhand von zwei Testdatensätzen miteinander verglichen. Dabei erhalten wir für die Detektion von Dachebenen Genauigkeiten von 95% bzw. 91%. Im Falle der Dachformen können wir 85% aller Dächer richtig zu einer der vordefinierten Dachformen zuordnen. Wir zeigen, dass unser System fähig ist, Dächer zu rekonstruieren und deren Aufbauten mit hoher Genauigkeit zu detektieren. Der Ausbaugrad eines Daches kann dabei mit einer Detektionsrate von 87% bestimmt werden.

Wir zeigen in dieser Arbeit, dass vertikale Luftbilder eine sehr nutzvolle und ökonomisch interessante Alternative zur Generierung und Interpretation von 3D Stadtmodellen sind. Unsere Experimente wurden anhand einer Vielzahl von Datensätzen ausgewertet und zeigen, dass die

von uns verwendeten Datenquellen und Algorithmen robust sind und eine hohe geometrische Genauigkeit liefern und daher für eine weitere Liegenschaftsbewertung verwendbar sind.

Contents

ACKNOWLEDGEMENT 5
ABSTRACT 7
KURZFASSUNG 10
Contents 13
LIST OF FIGURES 17
LIST OF TABLES 25
Conference Papers 27
Journals 28
1. **MOTIVATION** 29
1.1. Describing Real Properties 29
1.2. Evolving from Digital Geodata to Location-Awareness 30
1.3. Towards Semantically Interpreted Urban Models 32
1.4. Growing from Location-Awareness towards Ambient Intelligence 32
1.5. Valuating Real Property 33
1.6. Interest in Real Estate Valuation 34
1.7. Contributions 35
1.8. The structure of the work 37
2. **STATE-OF THE ART** 39
2.1. Introduction 39
2.2. 3D Building models generated from imagery 42
2.3. 3D Building models generated from LIDAR data 43
2.4. Building Models and 3D City Models 44
2.5. Building Facades 48
2.6. 3D Reconstruction 49
Structure from Motion and Aerial Triangulation 49
Dense Matching 50
Range Image Fusion 51
Variational Methods for 3D reconstruction 51
2.7. Real property evaluation using aerial images 52
3. **FROM CHARACTERIZATION TO VALUATION** 53
3.1. Some general definitions for Real Property Valuations 53
Comparative value method: 55

Gross rental method: 56
Asset value method: 56
3.2. What can aerial photos contribute? 56
3.3. From the characterization to a valuation 57
3.4. Proposed Processing Framework 58
3.5. Experimental Datasets 61
4. **SOURCES OF GEODATA** 65
4.1. Introduction 65
4.2. Vertical Aerial Images 67
History of Vertical Aerial Images 67
Redundancy 71
4.3. Satellite Imagery 72
History of Satellite Imagery 72
4.4. Oblique Aerial Images 74
History of Oblique Aerial Images 75
Revival of Oblique Imagery 75
4.5. Aerial Laser Scanning (LiDAR) 76
History of LiDAR 76
Working Principle 76
4.6. Street Level Images and Lasers 78
Working Principle 79
4.7. Crowd Sourced Image Data & Community Photo Collections 79
4.8. Unmanned Aerial Vehicles (MAVs) 80
4.9. Catastre 81
4.10. Optional Street Network and Address Data 82
4.11. Derived Products 83
Aerial Triangulation 83
DSM and DTM Generation 84
Classification into Segmented Images 86
True Orthophoto Generation 87
4.12. Experiments: LiDAR versus 3D Vision 87
4.13. Experiments: Oblique versus Vertical Photography 98
4.14. Discussion and Summary 105
5. **DATA PER PROPERTY** 107

5.1. Introduction 107
5.2. Coordinate Transformations 108
Datum 108
Coordinate Systems 110
Linear conformal coordinate Transformation 112
5.3. Image Matching 113
Intensity Based Matching 113
Relational Methods 114
Feature Based Matching 114
Chamfer Matching 117
5.4. Minimizing Building Overhangs and Mismatches 119
5.5. Building Façade Separation 120
Proposed Method 121
5.6. Data per Property 125
5.7. Experiments 125
Accuracy of Segmentation and the Cadastre 126
Matching the Cadastre with the True Orthophoto 127
Evaluation of Façade Separation Method 132
6. **CHARACTERIZATION OF REAL PROPERTIES** 135
6.1. Introduction 135
6.2. Building Characterization 136
Computing Footprints of Facades in 2D 136
Determination of Facade Quadrilaterals 140
Projecting the Façade Quadrilaterals into the Aerial Images 142
6.3. Experiments 142
7. **FAÇADE DETAIL CHARACTERIZATION** 145
7.1. Introduction 145
7.2. Hidden Façade Removal 146
7.3. Planar Facades 147
Number of Floors 148
Windows 150
7.4. Complex Facades 151
Multi View Dense Matching 152
Image Correlation 155

Depthmap Generation 155
Major Facade Plane Detection 157
7.5. Experiments 160
Hidden facades and elimination of "bad" façades 160
Analysis of Planar Façades 161
Analysis of the 3D Façade Point Clouds 163
Oblique versus Vertical Imagery 166
Discussion 169
7.6. Conclusion 169
8. **ROOF DETAIL CHARACTERIZATION** 171
8.1. Introduction 171
8.2. DSM Smoothing 172
8.3. Plane Detection 173
Classical RANSAC Approach 174
J-Linkage Approach 175
Multi Label Segmentation 176
8.4. Classification 177
Initial Classification 178
Superstructure Classification 179
Refined Classification 181
8.5. Skylight Detection 185
8.6. Roof Characterization 186
Roof Shapes 187
Rooftop Extensions 190
Suitability for the Installation of Solar Panels 191
8.7. Experiments 192
DSM Smoothing 193
Plane Detection 193
Classification 199
Skylight detection 204
Roof characterization 204
8.8. Discussion 211
9. **CONCLUSION AND OUTLOOK** 213

LIST OF FIGURES

FIGURE 1: TYPICAL 3D CONTENT IN SUPPORT OF AN INTERNET SEARCH, TAKEN FROM MICROSOFT'S WWW.BING.COM/MAPS. THE FOCUS IS ON PROVIDING A PLEASANT "USER EXPERIENCE". THE EXAMPLE IS THE COLORADO STATE CAPITOL. THE GEOMETRY IS FROM VERTICAL AERIAL PHOTOGRAPHY, THE PHOTO-TEXTURE FROM SEPARATE AERIAL OBLIQUE IMAGERY. THE VEGETATION HAS BEEN CLASSIFIED AND COMPUTER GENERATED TREES HAVE BEEN PLACED AT THE LOCATIONS CLASSIFIED AS TREES. THIS WEB PRESENTATION IS NO LONGER SUPPORTED BY MICROSOFT 32

FIGURE 2: PROPERTY VALUES IN THE USA. THESE "VALUES" ARE DERIVED FROM PROPERTY TAX RECORDS IN EACH MUNICIPALITY, AS WELL AS FROM PUBLIC RECORDS OF REAL ESTATE SALES. (SOURCE: WWW.ZILLOW.COM 34

FIGURE 3: 3D CITY MODELS GENERATED USING DIFFERENT LODS [KOLBE, 2006] 47

FIGURE 4: RANGE IMAGE (DEPTH MAP) OF AN AIRBORNE IMAGE OF GRAZ AND OBLIQUE VIEW OF A TEXTURIZED DEPTH MAP. DEPTH MAP AND TEXTURED 3D MODEL WERE CREATED USING INSTITUTES DENSE MATCHING PIPELINE. 51

FIGURE 5: DIAGRAM OF THE PROPOSED WORK FLOW TO CHARACTERIZE REAL PROPERTIES FROM AERIAL IMAGES AND ASSOCIATED CADASTRAL DATA. 59

FIGURE 6: TRUE ORTHOPHOTO OF THE CITY OF GRAZ, 400 M * 400 M, AT A GROUND SAMPLING DISTANCE OF 10 CM. "TRUE" IN THIS ORTHOPHOTO IMPLIES THAT THE AERIAL PHOTOGRAPH HAS BEEN RE-PROJECTED CONSIDERING THE ELEVATION OF THE BUILDINGS SO THAT ONE CANNOT SEE ANY OF THE BUILDING FACADES. SHOWN ARE 216 BUILDINGS DETAIL: BROCKMANNGASSE, KOPERNIKUSGASSE; ULTRACAMX, ACQUIRED IN 2007. (COURTESY MICROSOFT-GRAZ) 60

FIGURE 7: CADASTRAL MAP OF GRAZ WITH LEGAL PROPERTY BOUNDARIES IS SHOWN FOR THE AREA IN FIGURE 3. THE ORIGINAL DATA IS IN VECTOR FORMAT THAT HAS BEEN CONVERTED TO A RASTER IMAGE AT A GSD OF 10CM. SHOWN ARE 300 SEPARATE PARCELS [COURTESY BEV-AUSTRIA].....61

FIGURE 8: TRUE ORTHOPHOTO OF THE CITY OF GRAZ COVERING PART OF THE DENSE URBAN CORE OF THE CITY [KLUCKNER, 2011]. 62

FIGURE 9: ORTHOPHOTO SHOWING THE DOWNTOWN AREA OF THE CITY OF ANNECY, FRANCE FLOWN WITH A GSD OF 10CM [LABORATOIRE MATIS, IGN PARIS].....63

FIGURE 10: CADASTRAL BOUNDARIES FOR PART OF THE URBAN CORE OF GRAZ. GREEN RECTANGLES ILLUSTRATE THE FOUR TEST AREAS. .64

FIGURE 11: WORKING PRINCIPLE OF PUSHBROOM CAMERAS. "*CROSS-TRACK COVERAGE IS DETERMINED BY THE NUMBER OF DETECTORS IN THE FOCAL PLANE AND THE FIELD-OF VIEW (FOV) OF THE LENS SYSTEM. THE FORWARD MOTION OF THE PLATFORM PRODUCES ALONG TRACK COVERAGE.*" [ZHANG ET AL., 2005] 69

FIGURE 12 (LEFT) GROUND COVERAGE OF THE ULTRACAM SERIES; (RIGHT) PHOTOGRAPH SHOWING ULTRACAMX BY VEXCEL IMAGING .72

FIGURE 13: SATELLITE IMAGE OF ROOSEVELT ISLAND, NEW YORK (GEOEYE-1 SATELLITE GSD 50CM; 2009) [FROM WWW.GEOEYE.COM] 74

FIGURE 14: LIDAR WORKING PRINCIPLE [BC-CARMS, 2006].....77

FIGURE 15: EXAMPLE OF MOBILE MAPPING SYSTEM MOUNTED ON A CAR [TAKEN FROMWWW.NAVTEQ.COM] 78

Figure 16: Example generated sparse point cloud using community photo collection showing St Stephans cathedral and Kärntnerstrasse in Vienna [Irschara et al., 2007] 80

Figure 17: (left) example of an UAV; (middle) collected images of an atrium scene (right) computed sparse point cloud from these images [Leberl et al., 2012].....81

Figure 18: Street layer for the area in Figure 6 obtained from the website http://www.openstreetmap.org/, Nov-2011). The raster GSD has been resampled to fit the aerial imagery sources at a GSD of 10 cm. To the right is a "hybrid" view of the street layer superimposed onto the orthophoto. 83

Figure 19: Aerial Traingulation (AT) result from 3000 aerial images covering an area of approximately 150 km² of Graz and surroundings [Irschara et al., 2007].....84

Figure 20: Digital Surface model for one test area from Graz data set (Vexcel Imaging) 85

Figure 21: Classification Image for one test area from Graz data set (Vexcel Imaging) 86

Figure 22: Color aerial photograph taken with the UltraCamXp digital aerial camera, and superimposed the measurements from the Leica ALS50 laser scanner. Direction of flight is to the right. Laser sampling interval is at 40 cm, pixel size of the image is at 8 cm; taken from [Leberl et al., 2010a].....91

Figure 23: ALS laser scan to the left, photogrammetric 3D to the right. Sampling interval of the ALS at 45cm, of the aerial photography at 8 cm; taken from [Leberl et al., 2010a]. 91

Figure 24: Laser scan of random city streets obtained with the StreetMapper (UK); image taken from [Leberl et al., 2010a]. 92

Figure 25: 60 laser points/m2 and 6000 image pixels per m2; taken from [Leberl et al., 2010a]. 92

Figure 26: Scene for an automated point cloud from street side imagery with a pixel size of 1.3 cm; taken from [Leberl et al., 2010a]. 93

Figure 27: An automated triangulation connects 18 images with overlaps in the selected scene (Figure 26); taken from [Leberl et al., 2010a]. 94

Figure 28: left: Individual point cloud patch resulting from an image triplet and processed on a PC with one GPU, taking about 2 seconds. The facade is partly obscured by bicycles, moving people and a car; taken from [Leberl et al., 2010a]. right: Fusion result obtained from 16 separately computed patches of point clouds, each from its triplet; taken from [Leberl et al., 2010a]. 94

Figure 29: Two aerial images showing Herz Jesu Church in Graz; (left) oblique aerial image; Detail from Microsoft Bing Maps. (right) vertical aerial image 98

Figure 30: Projecting a rectangular image format onto a trapezoidal shape on the ground. The camera position is 850 off from the near range [from Prof. F. Leberl] 100

Figure 31: Detail from Microsoft Bing Maps. Left is the orthophoto and superimposed outline of an oblique aerial image produced with the Pictometry system operated by Blom. Right is the oblique aerial image. 101

FIGURE 32: OBLIQUE AERIAL IMAGE AT 45° LOOK ANGLE TAKEN FROM MICROSOFT BING MAPS (LEFT); VERTICAL AERIAL IMAGE OBTAINED FROM ULTRACAMX AT A LOOK ANGLE OF 27° (RIGHT). THE MARKED SECTIONS IN THE ORIGINAL IMAGE SEGMENTS (ABOVE) HAVE BEEN RECTIFIED (BELOW). 102

FIGURE 33: THE MARKED SECTIONS OF FIGURE 32 HAVE BEEN RECTIFIED. AT LEFT ARE TWO SECTIONS FROM THE OBLIQUE DATA AT 45°, AT RIGHT FROM THE VERTICAL DATA AT 27°. 102

FIGURE 34: SIDE LOOKING CAMERA WITH OFF NADIR ANGLES AND GSD IN "RANGE DIRECTION" (GSDR) AND PIXEL SIZE ON THE FACADE (FSDV). 104

FIGURE 35: DIFFERENT GRADIENT SECTORS FOR EDGE DETECTION 116

FIGURE 36: PROCESSING STEPS OF CHAMFER MATCHING ALGORITHM [BURGER ET. AL 2008] 118

FIGURE 37: OVERLAYING THE CADASTRAL MAP (HIGHLIGHTED IN RED) OVER THE TRUE ORTHOPHOTO SHOWS DISPLACEMENTS OF THE CADASTRAL BOUNDARIES OR MISSING CADASTRAL BOUNDARIES VIS-À-VIS THE PHOTOGRAPHY. IN THIS CASE THE CADASTRAL BOUNDARY BETWEEN THESE TWO BUILDINGS IS MISSING. 120

FIGURE 38: THE CLASSIFICATION LAYER "*BUILDING*" IS BASED ON COLOR, TEXTURE AND ELEVATION VALUES. (A) IS THE BINARY LAYER; (B) PRESENTS THE CONTOURS AS A RASTER IMAGE AND (C) ARE THE DETECTED FAÇADE FOOTPRINTS FOR THE AREA WITHIN THE GREEN RECTANGLE HIGHLIGHTED IN RED 121

FIGURE 39: ´VERTICAL AERIAL IMAGE SHOWING FAÇADE STRIP AND SUPERIMPOSED DETECTED FAÇADE POLYGON MARKED IN GREEN (A) PROJECTED FAÇADE QUADRILATERALS WITHOUT THE USE OF HEIGHT INFORMATION AND (B) PROJECTED FAÇADE POLYGON BY INCORPORATING HEIGHT INFORMATION. 122

FIGURE 40: ABOVE -- HEIGHT PROFILE OF A FAÇADE STRIP; BELOW -- DETERMINED FAÇADE SPLITS (HIGHTLIGHTED IN RED) BASED ON HEIGHT DISCONTINUITIES VISIBLE IN HEIGHT PROFILE 123

FIGURE 41: DETECTED HARRIS CORNERS AND THE EXTRACTED INTENSITY PROFILES CONSISTING OF 20 VALUES FOR EVERY RGB CHANNEL; TAKEN FROM [WENDEL ET AL., 2009] 124

FIGURE 42: FROM STREETSIDE DATA TO SEPARATION: (A) MATCHING OF ARBITRARY AREAS (B) DETECTED REPETITIVE PATTERNS (COLOR CODED LINES) (C) PROJECTION RESULTS IN A MATCH COUNT ALONG THE HORIZONTAL AXIS (D) THRESHOLDING THE REPETITION LIKELIHOOD WITH THE UNIFORM REPETITION LIKELIHOOD (E) RESULTING REPETITIVE AREAS, SEPARATION AREAS (GREEN), AND UNKNOWN AREAS (RED);TAKEN FROM [WENDEL ET AL., 2009]. 124

FIGURE 43: SAMPLE PROPERTY AND THE SOURCE DATA FOR THAT PROPERTY. (A) THE ORTHOPHOTO SEGMENT (B) DSM ELEVATION DATA (C) THE ELEVATIONS ON TOP OF THE BALD EARTH. (D) SHOWS THE CLASSIFICATION RESULT WITH SEVEN OBJECT CLASSES....125

FIGURE 44: EVALUATION OF BUILDING CLASSIFICATION (A) COMPUTED BUILDING CLASSIFICATION (B) MANUALLY DETERMINED GROUND TRUTH (C) DIFFERENCE IMAGE (RED: CORRECTLY DETECTED BUILDINGS, BLUE: FALSE POSITIVES, GREEN: FALSE NEGATIVES) 127

FIGURE 45: OVERLAYING THE CADASTRAL MAP OVER THE ORTHOPHOTO WILL LEAVE SOME SMALL ERRORS THAT NEED TO BE FOUND AND REMOVED. LEFT IS THE OVERLAY AFTER STEP 1 OF THE REGISTRATION, NAMELY THE COORDINATE TRANSFORMATION FROM THE GAUSS-KRÜGER SYSTEM TO THE UTM SYSTEM. TO THE RIGHT IS THE RESULT AFTER A CHAMFER- MATCH HAS BEEN APPLIED. 128

FIGURE 46: INPUT DATA TO THE CHAMFER MATCH. EDGE DETECTION RESULTS FOR (UPPER LEFT) TRUE ORTHOPHOTO (UPPER RIGHT) DSM (LOWER LEFT) BUILDING CLASSIFICATION AND (LOWER RIGHT) CALCULATED DISTANCE IMAGE FROM CADASTRAL DATA. 129

Figure 47: Detected Building and Roof overhangs (a) applying minimum area parameter (b) applying extent parameter and (c) combination of both results. 131
Figure 48: Overlay of building classification and cadastre. Left is with the discrepancies due to roof eaves and façade detail, to the right is the cleaned-up version....132
Figure 49: Result of façade separation (green and red lines are detected splits) (a) without height information, (b) including height information and (c) façade with false positive marked in red 133
Figure 50: The classification layer "building" is based on color and texture. The left shows the binary layer, center is its contour in raster and finally at right is the geometric figure of the footprint (single lines are marked in different colors). 136
Figure 51: Determined line segments based on the building classification. Colored lines are detected line segments. ...137
Figure 52: Determined line segments marked in different colors based on Vertical Elements in the DSM 138
Figure 53: Merging of line segments from the two datasets shown in Figure 51 and Figure 52....139
Figure 54: Building Footprints of two buildings from our test area. In these cases building classification is wrong and would lead to incorrect results when characterizing the facades....139
Figure 55: Building Footprint with two virtual facades (highlighted in blue and green) and footprint corrected footprint 140
Figure 56: Images shows 2D footprints and height profile along these footprints of two buildings (above) Simple footprint with constant height profile. (below) Complex footprint with height discontinuities along elevation profile (highlighted in red in the histogram); Red line illustrates façade split....141
Figure 57: (a) True orthophoto of Graz test area with example building highlighted in red and relevant façade footprint; (b) one vertical aerial image showing the respective building façade. (c-e) Corner points of one façade of one building projected in three different vertical aerial photographs. The façade quadrilateral is superimposed over the aerial photography to obtain the photo texture for visual inspection....143
Figure 58: Determination of planar and complex facades (white line shows building footprint derived from building classification; red line is simplified façade footprint; yellow area contains points that lie within 50cm of the footprint.)....146
Figure 59: Building facades extracted from different overlapping aerial images; (above) three façades that are visible in aerial images and thus can be further processed (same as in figure 60) (viewing angles: 25°, 26.2°, 19.7°) and (below) two images where façades are hidden and thus have to be eliminated....147
Figure 60: One building façade visible in four different vertical aerial images. These have been rectified into a façade coordinate system. From an aerial image block showing for each object point typically 10 images, not all will contain useful data for a specific vertical façade. Selected here are the 4 best, where "best" is defined as the largest area of a façade quadrilateral in the projection into an image. 148
Figure 61: Binary Prewitt edges in (a) are vertical, in (b) horizontal. The sums of edge values are shown in (c) as a count of the number of floors 149

Figure 62: Four images on one façade from Figure 60 lead to independent floor and window counts. It is to be noted that the floor counts and the number of windows (highlighted in blue) coincide with the visual inspection (see figure 63).....151

Figure 63: To the left is a rectified facade image with its depth structure. To the right is a failed count of windows using the workflow that assumes that a facade is in a single plane. Abandoning this restriction will cause a need for 3D point clouds of facades. 151

Figure 64: Roof Characterization Framework.....152

Figure 65: Definition of the two constraints angle-of-view a and direction of the façade b.....153

Figure 66: Plane sweeping principle. The homography between the façade's reference plane and the sensor view varies for different depths. [Szeliski et al., 1999] 153

Figure 67: Minimal and maximal depth range for one complex façade (blue: approximated façade footprint; green: maximal depth range; red: minimal depth range); 154

Figure 68: Correlation coefficients calculated for 4 different planes visualized as binary images (white areas have the largest correlation values).....155

Figure 69: (a) Computed depthmap of one façade (see figure 73a) and (b) overlaid with photo texture (left: $\delta = 10$) (right: $\delta = 2$).....156

Figure 70: Plane detection for a complex facade; (a) key-view of a facade and below projection in xy-plane (red lines are strongly weighted (>66% of points lie in this plane and green lines are balconies); (b) determined major façade planes (segmented areas with balconies are marked in red); (c) side-view (yz-plane) of turquoise façade plane (projection of points into yz-plane) 158

Figure 71: Enhancement of building classification. (a) Building Footprint of one building; (b) determined façade footprints; (c) overlaid enhanced building footprint (d) modified enhanced building footprint. 158

Figure 72: Plane detection using 3D approach; (a) key-view of facade; (b) approximated major façade planes without regularization; (c) labeled façade image.....159

Figure 73: Floor and Window detection for the determined labeled regions using the key view 160

Figure 74: Two examples for floor detection using edge histograms. (left) oblique aerial image (right) vertical aerial image; 166

Figure 75: Window detection approach using edge histograms. Images a and b illustrate detected windows using oblique and vertical aerial images; Images (c) and (d) show vertical and horizontal summarized and normalized edges ...168

Figure 76: Sequence of the five major processing steps 172

Figure 77: (a) RGB image of single building of the Graz test dataset; (b) Raw range image prior to smoothing; (c) Range image using TGV; (d) Detail of raw range image; (e) Smoothed detail. 173

Figure 78: Axonometric view of (a) point cloud of building roof, points in red; (b) point cloud and detected major planes in green. Note the intersection line and points belonging to the planes in black (c) point cloud and detected minor

planes in blue. Note intersection line and points belonging to the planes in black; (d) side view of point cloud and major planes; at right is an enlarged detail showing standard deviation between points and plane at ±12cm. 176
Figure 79: Proposed classification processing steps divided into two major steps. 178
Figure 80: Initial classification result for four buildings (green: major roof planes; blue: smaller roof planes; red: superstructures).....179
Figure 81: Requirements for chimneys on building roofs; taken from.....180
Figure 82: Different types of dormer windows [taken from Wikipedia] 181
Figure 83: Superstructure classification for four buildings based on initial classification from Figure 80. yellow: chimneys; red: dormer windows; green: other structures 181
Figure 84: The seed pixel and its 8 surrounding neighbors that are used for region growing 183
Figure 85: Contour of one extracted chimney; Illustrated are different steps of bounding fitting rectangle: (a) shape of chimney after region growing; (b) centroid and board of the chimney; (c) major and minor axes of area; (d) upper and lower furthest points; and (e) four vertices of bounding rectangle.....184
Figure 86: Classification of a test buildings; (a) major roof planes (in red); (b) smaller roof planes (in yellow); (c) dormer windows (in blue); (d) chimneys (in green); 185
Figure 87: Refined Superstructure Classification after Rectangle Fitting for four example buildings.....185
Figure 88: Left: original roof image; Right: detected skylights (marked in green) on a building roof.....186
Figure 89: (a) building visible in vertical aerial image (b) segmentation result (c) in red modified building outline using 3D façade reconstruction (d) enhanced building footprint. 187
Figure 90: Decomposition of building footprint from figure into quadrilateral cells. 188
Figure 91: Major architectural roof styles (a) gable roof (b) shed roof (c) hip roof building (d) half hipped roof (e) mansard roof (f) flat roof (g) saw-tooth roof (h) mansard hipped roof. From [www.projekt--service.de/dacharten.html].....189
Figure 92: Classification of a building roof into different height classes; Shown are detected lower and upper class for one building roof. taken from [Dan, 1996].....189
Figure 93. Examples for possible connecting roof shapes with normal vector 190
Figure 94: Compilation of different features for classification into the roof styles 190
Figure 95: (left) classified roof image; (right) interpreted classification (extended attic is highlighted in pink) 191
Figure 96: Annual sum of incoming global solar radiation on roof planes in kWh/m2 by considering transparent properties of vegetation; image taken from [Jochem et al., 2009].....191
Figure 97: Ground Truth for Graz data set (projected on xy-plane) (red: major planes; yellow: smaller planes; blue: dormer windows; green: chimneys) This was collected manually.....192
Figure 98: Ground Truth for Graz data set (oblique viewing direction) 193
Figure 99: Results for segmentation by variation of the regularization parameters Λ and Γ. Within a reasonable range, the segmentation result is very robust. values for Λ ranging from 0 to 0.15 and Γ ranging from 500 to 1250.....197

Figure 100: Results for height model segmentation using fixed parameters (λ = 700 and Γ = 0.05) (a) Input depth map, segmentation image, input rendering, reconstructed building.....198
Figure 101: Different examples of errors within the height model segmentation.....199
Figure 102: Building footprints based on classification image and after incorporating 3D façade information.....206
Figure 103: Overlay of building footprints from Figure 102. Detail of overlay shows discrepancies at building edges due to errors within the image classification. Building footprint before enhancement (red) and derived footprint after 3D reconstruction of building facades (green).....206
Figure 104: Cell Decomposition. Buildings are splited into rectangular areas for further processing.....207
Figure 105: Examples for non allocatable building roofs (Graz data set showing Graz Opera House and extended rooftop of modern residential building) 208
Figure 106: Ground truth for Graz test data set. Yellow: Buildings with dormer windows; Red: Buildings with skylights; Green: Buildings with dormer windows and skylights.....210
Figure 107: Results of Rooftop Analysis. (left) Ground Truth for extended building rooftops. Extended rooftops are highlighted in pink; (right) Result of Rooftop Analysis. (green) correctly detected extended rooftops (red) false negatives (blue) false positives 211

LIST OF TABLES

TABLE 1: PRICE DIFFERENCES IN VIENNA DEPENDING JUST ON THE LOCATION OF THE PROPERTY (SOURCE: HTTP://IMMOMEDIA.KURIER.AT, ACCESSED 13 JANUARY 2012) 35
TABLE 2: DESCRIPTION OF DIFFERENT LEVELS OF DETAIL INCLUDING THEIR ACCURACIES [KOLBE, 2006] 46
TABLE 3: DIFFERENT VALUES IN CONNECTION TO REAL PROPERTY AND THEIR RELEVANCE [METZGER, 2006] 55
TABLE 4: DETERMINATION OF VALUE AFFECTING FACTORS DERIVED FROM AERIAL IMAGES 57
TABLE 5: THE MAJOR SOURCES OF URBAN SOURCES FOR GEO-DATA AND THEIR EXPECTED TYPICAL GEOMETRIC RESOLUTION. A CONSIDERABLE DYNAMIC EXISTS TO CONFLATE IMAGE, VECTOR AND ADDRESS DATA, TO CREATE UPDATING MECHANISMS FOR STREET-LEVEL AND ADDRESS DATA [GOOGLE BLOG ON CONFLATION AND UPDATING, [1, BLUMENTHAL, 2009]. 66
TABLE 6: LIST OF ADVANTAGES OF DIGITAL SENSING OVER FILM CAMERAS IN PHOTOGRAMMETRY [LEBERL ET AL., 2004] 70
TABLE 7: TECHNICAL PARAMETERS OF MOST COMMON LARGE FORMAT DIGITAL AERIAL CAMERAS [JACOBSEN, 2011] 71
TABLE 8: CHARACTERISTICS OF MAIN SENSORS FOR SATELLITE IMAGERY [PETRIE, 1999] 74
TABLE 9: PROJECT PARAMETERS FOR A LASER SYSTEM AND A DIGITAL PHOTOGRAMMETRIC CAMERA. NOTE THE DIFFERENCES IN STRIP WIDTH AND AIRCRAFT SPEED, WHICH CAUSE THE PHOTOGRAMMETRIC DATA COLLECTION TO BE ALMOST 13 TIMES MORE ECONOMICAL AT THESE PARAMETERS. IN ADDITION, THE HIGHER NUMBER OF PHOTOGRAMMETRIC POINTS PER M2 PRODUCES A BETTER DEFINITION OF THE OBJECT SHAPE ALONG SHARP EDGES. AND THE HIGHER SIDE-LAP OF THE IMAGE-BASED METHOD REDUCES THE OCCURRENCE OF OCCLUSIONS. [LEBERL ET AL., 2010A] 89
TABLE 10: ERROR VALUES OF THE STREET-SI DE LASER DATA IN CENTIMETERS, TAKING ADVANTAGE OF THE 4 PASSES OF THE KRONES SCHOOL. THE "DELTA-VALUES" ARE THE R.M.S. DEVIATIONS OF THE POINTS FROM A PLANE IN EACH PASS SEPARATELY SINCE THE POINTS ARE OF A VERTICAL WALL. THE "DEVIATIONS OF THE MEAN" ARE THE DIFFERENCES OF EACH POINT CLOUD'S MEAN FROM THE OVERALL MEAN FORMED FROM ALL POINTS IN ALL 4 POINT CLOUDS. WE FIND A NOISE LEVEL IN THE 2-CM-RANGE; TAKEN FROM [LEBERL ET AL., 2010A]. 95
TABLE 11: SIXTEEN ADVANTAGES OF THE PHOTOGRAMMETRIC 3D WORKFLOW OVER THE DIRECTLY MEASURED LASER POINT CLOUDS; TAKEN FROM [LEBERL ET AL., 2010A] 96
TABLE 12: IMAGE PARAMETERS DERIVED FROM OBLIQUE AERIAL IMAGES EXTRACTED FROM BINGMAPS WEBSITE [FROM PROF. F. LEBERL] 99
TABLE 13: BING/MAPS OBLIQUE IMAGERY PARAMETERS RECONSTRUCTED FROM KNOWN TERRAIN POINTS. 100
TABLE 14: SOME GEOMETRIC DATA OF TWO TYPICAL DIGITAL AERIAL CAMERAS (FROM WWW.VEXCEL.COM) 101
TABLE 15: INCIDENCE OR LOOK ANGLES AND VERTICAL PIXEL SIZE WITHIN A FACADE. THE HORIZONTAL PIXEL SIZE IS AT 10 CM. 103
TABLE 16: SIZE OF A PIXEL ON A FACADE IN CM, AS A FUNCTION OF THE LOOK ANGLE, IN DEGREES 104
TABLE 17: DATUM FOR THE MOST COMMON REFERENCE ELLIPSOIDS [HTTP://EN.WIKIPEDIA.ORG/WIKI/REFERENCE_ELLIPSOID] 109
TABLE 18: STANDARD PARAMETERS FOR COORDINATE TRANSFORMATION WGS84 – GK (SOURCE: BEV 2011) 113
TABLE 19: EVALUATION OF BUILDING CLASSIFICATION BASED ON PIXELS 126
TABLE 20: EVALUATION OF COORDINATE TRANSFORMATION; MISMATCHES BETWEEN CADASTRE AND ORTHOPHOTO IN PIXELS (GSD 10CM) 128
TABLE 21: EVALUATION OF CHAMFER MATCHING: OFFSETS BEFORE AND AFTER CHAMFER MATCHING USING DIFFERENT INPUT DATA (IN PIXEL) 130
TABLE 22: PROCEDURE FOR FLOOR DETECTION IN SIMPLE PLANAR FACADES 149
TABLE 23: FAÇADE IMAGES CLASSIFIED INTO DIFFERENT VIEWING ANGLE CATEGORIES BEFORE INCORPORATING REJECTION THRESHOLDS AND RESULTS AFTER INCORPORATING REJECTION THRESHOLDS. 161
TABLE 24: CLASSIFICATION OF THE BUILDING FACADES INTO DIFFERENT CATEGORIES DEPENDING ON THE NUMBER OF IMAGES THAT ARE VISIBLE. 161
TABLE 25: DIVISION OF THE FACADES INTO COMPLEX AND PLANAR FACADES 161
TABLE 26: PLANAR FAÇADES CLASSIFIED INTO DIFFERENT VIEWING ANGLE CATEGORIES. 162

Table 27: Planar Façades classified into different categories depending on the number of images that these facades are visible in. 162
Table 28: Counting floors (above) and windows (below) from vertical images versus look angles. 162
Table 29: Complex Façades classified into different viewing angle categories..... 163
Table 30: Evaluation of the facade plane detection for planar and complex facades using the column-wise approach 164
Table 31: Evaluation of the facade plane detection for planar and complex facades using the 3D approach 164
Table 32: Assessment of the geometric quality of the point cloud 165
Table 33: Evaluation of Floor and Window Detection (planar and complex facades)..... 165
Table 34: Counting floors (above) and windows (below) from vertical images and results depending on look angles. Last column is from oblique images where floor counts are compromised by occlusions. 167
Table 35: Ground truth of roof planes and roof superstructures for Graz and Annecy data set 194
Table 36: Plane detection results using classical RANSAC approach 194
Table 37: Plane detection results using J-Linkage approach..... 195
Table 38: Plane detection results using multi label segmentation approach..... 196
Table 39: Ground truth for different types of Superstructures (Dormers, Chimneys and other structures)..... 200
Table 40: Gournd Truth for Superstructures (Dormers, Chimneys and other structures)for CORRECTLY detected superstructures ... 200
Table 41: Experimental results for the building roof segmentation showing the detected roof segments (Graz data set)..... 201
Table 42: Experimental results for the building roof segmentation showing the detected roof segments (Annecy data set) 201
Table 43: Misclassification of superstructures (Graz data set) 201
Table 44: Misclassification of superstructures (Annecy data set)..... 202
Table 45: Superstructure Detection: False positives..... 202
Table 46: Accuracy analysis of classification step before region growing (Graz data set) 203
Table 47: Accuracy analysis of classification step before region growing (Annecy data set) 203
Table 48: Accuracy analysis of classification step after region growing (Annecy data set) 203
Table 49: Accuracy analysis of classification step after region growing (Annecy data set) 203
Table 50: Detection rates for skylight detection for Graz data set 204
Table 51; Ground truth: types of roof shapes for the 216-building test area 205
Table 52: Detection rate for roof planes and superstructures after downsampling the range data by a factor four. 205
Table 53: Evaluation of roof shapes based on decomposed cells (Result including all cells) 207
Table 54. Evaluation of roof shape detection for Graz data set 209
Table 55: Evaluation of roof shape detection for Annecy data set..... 209
Table 56: Ground truth for 69 extended attics in 216 buildings. Some attics have only dormer windows or skylights, and some have both. 210

DOCTORAL RESEARCH: PUBLICATIONS

The following publications resulted from my doctoral research at the Institute for Computer Vision and Graphics at Graz University of Technology. They are sorted by date. The thesis builds on the material from these publications.

Conference Papers

[1] P. Meixner and F. Leberl 2010. "Framework to automatically characterize real property using high resolution aerial images", In Proceedings of the *Annual Convention of the* American Society for Photogrammetry and Remote Sensing ASPRS*, San* Diego, USA, Unpaginated CD.

[2] P. Meixner and F. Leberl 2010. "From Aerial Images to a Description of Real Properties: A Framework." Proceedings of VISAPP International Conference on Computer Vision and Theory and Applications, Angers.

[3] P. Meixner and F. Leberl 2010. „Vertical- or Oblique Imagery for Semantic Building Interpretation". Dreiländertagung OVG, DGPF und SGPF; Vol. 19, pp. 247-256.

[4] P. Meixner and F. Leberl2010. "Describing Buildings by 3-dimensional Details found in Aerial Photography". ISPRS Vol. XXXVIII, Part 3B, pp. 98-103.

[5] P. Meixner and F. Leberl 2010. "Characterizing Building Facades From Vertical Aerial Images. "In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 38, Part 3B, pp. 98-103. (**Best Poster Award**)

[6] P. Meixner and F. Leberl 2010. „Augmented Internet Maps with Property Information from Aerial Imagery", in 18th ACM-SIGSPATIAL Intl. Conf. on Advances in Geographic Information Systems, San Jose, November 3-5, Proceedings available from ACM, p. 406-409. (**Best Forward Presentation Award**)

[7] P. Meixner and F. Leberl 2010. „Interpreting Building Facades from Vertical Aerial Images Using the Third Dimension", in Symposium of Commission IV of the International Society for Photogrammetry and Remote Sensing [ISPRS], Florida, 2010, M. Madden (ed.): Archives of the ISPRS., vol. 38, Part 4

[8] P. Meixner and F. Leberl 2011. "Detecting Occlusions in Facade Interpretation from Vertical Aerial Images", in *Annual Convention of the* American Society for Photogrammetry and Remote Sensing ASPRS*, Milwaukee*, USA, Unpaginated CD.

[9] P. Meixner, F. Leberl and M. Brédif 2011. "3D Roof Details by 3D Aerial Vision", In: Proceedings IEEE Computer Society International Conference, In Computer Vision / *ISPRS Workshop on Computer Vision for Remote Sensing of the Environment CVRS11, 7-Nov-2011*.

[10] P. Meixner, F. Leberl and M. Brédif 2011. "Planar Roof Segments from Aerial Photogrammetry", *19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* . pp. 9-15.

[11] P. Meixner, F. Leberl and M. Brédif 2011. "Interpretation of 2D and 3D Building Details on Facades and Roofs", *PIA 2011 Conference on Photogrammetric Image Analysis, Munich, October 5-7;* In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 38, part 3/W22, pp. 137-142. (**Best Poster Award**)

[12] P. Meixner, A. Wendel, H. Bischof and F. Leberl 2012. „Building Facade Separation In Vertical Aerial Images". To appear in: *Proceedings of the 22th Congress of the ISPRS*, Melbourne, Australia.

Journals

[13] F. Leberl, A. Irschara, T. Pock, P. Meixner, M. Gruber, S. Scholz and A. Wiechert 2010.„Point Clouds: LIDAR versus 3D Vision", In: *Photogrammetric Engineering and Remote Sensing*, vol. 76, pp. 1123-113.

[14] P. Meixner and F. Leberl 2011. "3-Dimensional Building Details from Aerial Photography for Internet Maps", *Remote Sensing*, vol. 3, pp. 721-751.

[15] F. Leberl, P. Meixner, A. Wendel, and A. Irschara 2012, "Urban 3D Modeling using Aerial Photography and 3D Vision", *Optical Engineering, 51.02117, March 01,2012, http://dx.doi.org/10.1117/1.OE.51.2.02/117 , 2/2012, 12 pages*.

CHAPTER ONE

MOTIVATION

Contents

1.1	Describing Real Properties	29
1.2	Evolving from Digital Geodata to Location-Awareness	30
1.3	Towards Semantically Interpreted Urban Models	32
1.4	Growing from Location-Awareness towards Ambient Intelligence	32
1.5	Valuating Real Property	33
1.6	Interest in Real Estate Valuation	34
1.7	Contributions	35
1.8	The structure of the work	37

1.1. Describing Real Properties

We define a "real property" by one or more parcels as they are recorded in cadastral maps. It consists of a piece of land, sometimes bounded by a fence, on that land are one or more buildings, impervious surfaces, garages, trees and other vegetation. A property may also contain only the portion of a multi-unit building, for example in dense urban cores where buildings are connected.

The description of a real property consists of tables with coordinates and numbers. These define the number of buildings, type of building from a stored list of candidates, building height and footprint, number of floors, number and types of windows, presence of a basement floor, types of attics, roof types and roof details such as an eave, skylights, chimneys, presence of a garage and its size, types and extent of impervious surfaces such as driveways and parking spaces, and statements about the type and size of elements of vegetation, the presence of a water body, the existence and type of a fence, exposure to the sun and effects of shadows, the quality of views from a window etc.

We expect that such information per parcel, for all properties of a city, will be useful for a broad range of applications in city administrations, commercial location-based services, and searches on the Internet [Gartner et al., 2008]. There will, however, be concern about the cost of data creation. An economically favorable approach would build first on the wealth of already existing routine aerial photography justified by other applications, not insignificantly in

connection with innovative location-aware global Internet sites such as Google Maps and Microsoft Bing-Maps or the innumerable regional sites. And secondly, such property-descriptions will need to be developed automatically and with little manual support.

Therefore we have designed a processing framework with a focus on using pre-existing high resolution vertical aerial images and cadastral data. These two sources need to be conflated into a common coordinate system. The aerial images themselves in turn are the source of terrain models, image camera poses and classifications. Theoretically, one may also involve street vectors. The process first builds an integrated geometry and semantic object data set to be then analyzed for the development of descriptors.

1.2. Evolving from Digital Geodata to Location-Awareness

A location-aware Internet [Leberl et al., 2007; Leberl et al., 2009a; Leberl et al., 2009b; Leberl et al., 2003b] has evolved since 2005. Internet-search has been a driving force in the rapid development of detailed 2-dimensional maps and also 3-dimensional urban models. "Internet maps" in this context consist of the street-maps used for car navigation, augmented by the terrain shape in the form of the Bald Earth and all this being accompanied by photographic texture from orthophotos. This is what is available for large areas of the industrialized world when calling up the websites maps.google.com or www.bing.com/maps, and also from a number of regional Internet mapping services, for example the French www.geoportail.fr or the German www.klicktel.de/kartensuche.

The transition from the first availability of digital map data to today's interactive location-based search on the Internet spans around 30 years. One may divide the evolution into several distinct stages [14].

1. The massive transition to digital Geodata of course dates back to the time when computing facilities were first able to cope with the data quantity implied in the geo-application and began with the scanning of paper maps and their subsequent vectorization in the late 1970's. This improved the storage, retrieval, transmission and printing of such data.

2. The availability of digital data led to a start of an early form of pre-Internet e-commerce by shipping shrink-wrapped Geodata to interested experts.

3. Road maps became available for trip planning in the mid 1990's, initially for transportation systems, later for the general public in the form of car navigation. It required the broad acceptance of personal computing. Geodata were transported on digital media.

4. The Internet of course was the major enabler of the widespread commerce with Geodata to greatly improve the convenience in the exchange of orthophotos, aerial and satellite photography. Shrink-wraps were then replaced by online services.

5. The GPS inspired a further increase in the usefulness of digital Geodata and resulted in the transition to true real-time navigation, from about the mid 1990's onwards, based on street maps.

6. The latest stage of this evolution was the introduction of Geodata into the novel search-application via the Internet. Global location awareness was introduced into Internet search via maps.google.com and Microsoft's www.bing.com/maps. 3D urban building models came into being in November 2006 with Microsoft's announcement of the availability of Virtual Earth in 3D [Paul, 2006].

7. We see today the emergence of smart telephones and a refined interest in locations not only in cars, but also for pedestrian applications.

An interesting transition occurred along the way from Geodata as a realm of experts into routine use by everyone. This separate evolutionary track began with a very early idea, namely the invention of the GIS in about 1962 [Tomlinson et al., 1967]. The idea was born well ahead of the ability to realize it in the form of large area systems. While of course entirely predicated on the availability of the computer, it took the personal computer to make the idea a practical reality. However, the GIS did not change the basic limitations of "Geodata for the experts". That change became very visible with the development of car navigation systems for regular passenger cars. It signaled for the first time a transition from experts to everyone. This development accelerated with the location-aware Internet search so that we have today almost every Internet user also a user of Geodata, inspecting and judging aerial photography, oblique imagery, 3D building models and, most recently, street-side images and their various methods of supporting a Geo-naïve user.

Urban 3D models have been a research topic since the mid 1990's [Dang et al., 1993; Grün et al., 1995; Förstner et al., 1995; Gruber et al., 1995; Gruber, 2007]. The initial introduction into Internet-search at Bing/Maps (then Virtual Earth) employed man-made buildings as triangulated point clouds with a visual embellishment by photographic texture. Since April 2008, vegetation is being classified and identified, and computer-generated vegetation is being placed on top of the Bald Earth [Leberl et al., 2009b].

3D urban models still are in their infancy and are provided over large areas only by the Microsoft-web site Bing/Maps, with an example presented in Figure 1in its "classic" version from www.bing.com/maps. Recently, the emphasis has shifted at Microsoft from a full 3D urban model to the use of 2D street-side imagery that has undergone some limited 3D processing (see www.bing.com/maps/explore). While Internet-search may be the most visible and initial driving application, there are of course others. Often mentioned are city planning, virtual tourism, disaster preparedness, military or police training and decision making [Willkomm et al., 2009], or 3D car navigation [Strassenburg-Kleciak et al., 2007]. In Germany, 3D building models are being considered of such municipal relevance that a fairly complex and all-encompassing CityGML data standard has been developed to guide the creation of 3D building models at various levels of detail ranging from a basic L0 through a virtual reality level L4 with indoor 3D modeling [Gröger et al., 2007]. This standard has been adopted by the international OGC Open Geospatial Consortium [Kolbe et al., 2009].

Figure 1: Typical 3D content in support of an Internet search, taken from Microsoft's www.bing.com/maps. The focus is on providing a pleasant "user experience". The example is the Colorado State Capitol. The geometry is from vertical aerial photography, the photo-texture from separate aerial oblique imagery. The vegetation has been classified and computer generated trees have been placed at the locations classified as trees. This web presentation is no longer supported by Microsoft

1.3. Towards Semantically Interpreted Urban Models

So-called 3D-location awareness of the Internet serves to please the user's eye – one could speak of "eye candy" -- but cannot be used as part of the search itself. This is unlike the 2D content with its street map and address codes that can be searched. An interpreted urban 3D model would support searches in the geometry data, not just in the alphanumeric data. One may be interested in the number of buildings higher than 4 floors in a certain district, or in properties with specific characteristics such as: a built-up floor area in excess of 500 m2, impervious areas at less than 30% of the land area, large window surfaces in excess of 30% of the wall surface, a certain minimum sun exposure during the day, or attractive views of the distant urban landscape.

We can search for such objects if we augment the Internet with descriptions of real properties. Image contents must not merely be shown on a monitor, but they must also be interpreted. If this is accomplished at a low cost, it must be done automatically and from easily available sources. We are facing a research challenge for computer vision [Leberl et al., 2009a; Leberl et al., 2010b].

1.4. Growing from Location-Awareness towards Ambient Intelligence

While applications like Bing-Maps or Google Earth are currently driven to pull the public into using the associated Internet-search system via attractive location awareness, there is a deeper

justification in light of the emerging opportunities created by the Internet-of-Things and Ambient Intelligence. Both concepts have need for location awareness [O'Reilly et al., 2009]. In 1991, Marc Weiser authored his much quoted prediction for computing in the 21st century and postulated that location will be one of two issues of crucial importance: "ubiquitous computers must know where they are" [Weiser, 1991].

Weiser's sketch of the future has morphed from ubiquitous computing to ambient intelligence. We expect the geometry of the elementary parts of an entire urban environment to be available at a human scale to be searched; we foresee that this becomes the basis for locations of fixed sensors, or moving GPS-tracked sensors, to read the RFID- and other tags of goods to place these inside buildings, even inside individual rooms. In analogy, triangulated cell phones as well as tags can also place persons inside buildings and rooms. A semantically interpreted 3D city model could help find one's briefcase, an errant person or the nearest copying machine; thereby, affecting the use and behavior of a wide array of computer-driven gadgets.

1.5. Valuating Real Property

Real Estate valuation is an important topic in our daily lives. It forms the basis for economic action, because without it mortgages and property transactions would not be possible. During the last years, the Internet plays a larger role in connection with property assessment. The success of both Internet platforms "Google Earth" and "Virtual Earth" has shown that a big public interest exists in connection with interactively processed Geodata. Many experts say that the Internet is considered to be one of the biggest changes in the information technology with big effects on various areas of the everyday life. For example, one can access almost unlimited amounts of desired data from everywhere in the world at any time.

An example in the field of property assessment is the US homepage www.zillow.com, with which one can retrieve the value of every property and building in the US (see Figure 2). At www.zillow.com the value of a building will be derived from the property tax records of a municipality.

Valuation of real property should occur on the basis of externally recognizable factors and by the use of different databases. In the scope of this we want to determine parameters that enable a valuation of a building. In contrast to www.zillow.com, these building characteristics will be obtained from digital aerial images and cadastral records. The focus is to be on real property in Austria, with a view on other legal systems, for example in Eastern Europe which has a poor history of private land ownership.

Figure 2: Property values in the USA. these "values" are derived from property tax records in each municipality, as well as from public records of real estate sales. (Source: www.zillow.com

1.6. Interest in Real Estate Valuation

Such a property valuation system could be used for the calculation of mortgages. In case of the present financial crisis, banks and supervisory authorities could check mortgage loans, for example if the awarded amounts are justified. The supervisory authorities could determine fast and efficiently the true value of properties according to objective criteria and decide on whether it is still economically affordable to give a loan.

In many countries like USA, Great Britain, Canada and Australia the valuation of real property is done using the allocation of mortgage loans. This always happens individually for every building. A simultaneous valuation of all buildings in an area would be more efficient and comprehensible for everyone. One will not need thousands of people evaluating properties according to many different standards, but rather just one program that evaluates all the properties according to one standard simultaneously.

A traditional turn of phrase used by property experts is that the three most important factors in determining the desirability of a property are "location, location, location". You can buy the right home in the wrong location. You can change the structure, remodel it or alter the home's layout but, ordinarily, you cannot move it. It's attached to the land and its value strongly depends on the neighborhood where it is located. Table 1 shows an example that points out what big influence the location of a property has on the price.

Description of the building	District	m²	Price (€)
Cottage- mansion, 460m² living space, cellar, total parcel area 875m²	1190	460	2.900.000
Vacant residential building with 5 condominiums	1230	460	1.180.000
UNO-City Bruckhaufen with 650 m² living space, beautiful garden, 12 rooms, 4 bathrooms, 4 toilets	1210	450	1.000.000

Description of the land parcel	District	m²	Price (€)
Fantastic building site in top location	1130	1380	1.450.000
Bug parcel in Aspern	1220	1451	565.890
Building site - divisible	1220	1425	450.000
Building site with family home Vienna 22.,Donaustadt. Near Breitenleerstraße	1220	1598	290.000

Table 1: Price differences in Vienna depending just on the location of the property (Source: http://immomedia.kurier.at, accessed 13 January 2012)

Within the scope of the work, one should primarily examine which value-influencing factors can be won automatically from aerial images. The work will not go into detail on the location of a property at that time, because the relevant parameters for it can be obtained from external databases and would exceed the scope of this work.

1.7. Contributions

This reserach addresses the following research questions:

- Are Vertical Aerial images suitable for real property characterization and valuation?
- Do we need cadastral data and can we perform the characterization without it?
- How can we develop a framework to automate real property characterization?
- How can we combine the different data sets in an efficient way?
- How can we characterize building facades and which details can be extracted?
- How can we characterize buildings roofs and which details can be extracted?

- At what accuracy and completeness can we extract these details out of aerial images?

By means of the following tasks we try to clarify performance, possibilities and alternatives.

- Comparison of vertical aerial images with other kinds of source data.

  The characterization of real properties can be done using different kinds of source data. In our case we differentiate between two main data sources: airborne data and terrestrial data. Depending on the used technique, real properties can be characterized at different levels of detail and varying completeness. We compare the different airborne data sources for our purpose and demonstrate that for the interpretation of real properties vertical aerial images can compete with traditional LiDAR and oblique aerial images.

- Determine how the data has to be handled.

  A series of products can be derived from vertical aerial images like DSM, DTM, classification images and true orthophotos. We give an overview of how these products have to be combined to produce meaningful and accurate results and what are the bottlenecks of the used data.

- Analysis of the cadastral information

  We find discrepancies within the range of one meter that we can automatically remove, so that the residuals reduce to half a pixel. Cadastral information is important to divide land parcels and buildings attached to these parcels. We discuss the accuracy and completeness of cadastral data and the problems that occur when combining this data with airborne data. We also list alternative data sources for this purpose if cadastral information is not available.

- Characterization of building facades

  Depending on their structure facades are split into two categories: planar facades consisting of one plane and complex facades consisting of several planes with balconies and other building overhangs. There exists a series of methods for facade reconstruction and the extraction of its details. However, all these approaches don't use vertical aerial images. We show which methods are suitable for facades extracted from vertical aerial images and which façade details can be extracted. We reconstruct for the first time building facades in 3D from vertical aerial images and interpret the point cloud. This novel method is based on a total generalized variation multi-labeling approach (TGV).

- Characterization of building roofs

  Building roofs are a crucial issue for 3D city model generation. There exists a range of methods that deal with the problem of roof surface detection and reconstruction. However, depending on the used source data and its resolution, the result of this roof characterization can vary strongly. We discuss some of these methods and compare them. In addition, we introduce a novel approach of characterizing and reconstructing building roofs that is based on minimizing labeling costs.

- Improvement of building footprint

  Building footprints derived from aerial images do not cope with the footprints determined by terrestrial surveys. This is due to building and roof overhangs that affect the shape of the building contour. We propose methods to eliminate these overhangs fully and automatically without any human interaction resulting in an exact footprint that coincides with the real world. We show that we can successfully remove these overhangs.

- Accuracy and completeness of proposed methods

  Vertical aerial images are on first sight not suited for facade characterization. Therefore we have evaluated the used methods and determined their accuracy and completeness and show that it is feasible.

1.8. The structure of the work

The thesis is structured in the following way. **Chapter 2** introduces the theoretical background and the state of the art, beginning with the first projects in the field of image understanding. This is followed by an historic overview about 3D city model generation from aerial images and LiDAR data that build the basis for a characterization of these models. After that, the state-of the art methods for facade characterization and reconstruction are outlined.

Chapter 3 is structured in two parts. The first part is dedicated to general aspects of real estate valuation and which parameters can be extracted from images. The second part gives a short overview about the processing steps that are necessary to characterize real properties. The following chapters are dedicated to the single processing steps of this framework. This chapter contains material from [2].

Chapter 4 explains and lists the different kinds of source data and gives and overview about their development over the last decades. In the last part of chapter 4, the three major airborne source data sets are analyzed and a comparison is performed, which is heavily based on publications [3] and [13].

Chapter 5 presents the processing step data per property. Within this chapter the necessary coordinate transformation is described in more detail and the additional processing steps like chamfer matching and building and roof overhang removal. Finally, a novel facade separation method for vertical aerial images is introduced based on [11] and [12] followed by an evaluation of all used methods in this chapter.

Chapter 6 discusses the characterization of the real properties. It describes the processing steps for the determination of the 2D facade footprints for a building and afterwards the determination of the facade quadrilaterals and their projection into the aerial images. This is followed by an evaluation of the applied methods.

Chapter 7 is dedicated to the characterization of facade details like floors and windows. Firstly, it introduces a method to extract floors and windows from a planar facade. Secondly, complex facades consisting of building overhangs, several facade planes and balconies are

characterized. The results for both kinds of facades are presented also in this chapter. Finally, a comparison between oblique and vertical aerial images concerning the extraction of facade elements is performed. Material from this chapter can be found in the publications [3], [5] and [14].

In **Chapter 8,** the necessary processing steps for the characterization of building roofs are presented. In the first part a method to smooth the elevation data is presented. In the second part different methods to detect roof planes are explained. In the third part the planes are classified into different roof structure categories. Additionally, a method to detect skylights on roofs is presented. The last part is dedicated to roof characterization. We explain a method to determine the roof shape of any arbitrary building and how to determine the degree of extension of a building roof. This is followed by a detailed analysis of the implemented methods. This chapter is heavily based on [9] and [10].

Finally, **Chapter 9** concludes the work by answering the seven questions raised in the introduction. Further suggestions are given for future works.

CHAPTER TWO

STATE-OF THE ART

Contents

2.1	Introduction	39
2.2	3D Building models generated from imagery	42
2.3	3D Building models generated from LIDAR data	43
2.4	Building Models and 3D City Models	44
2.5	Building Facades	48
2.6	3D Reconstruction	49
2.7	Real property evaluation using aerial images	52

2.1. Introduction

The literature concerning the topic "Virtual City ", sometimes called "Virtual Habitats" and "Virtual Earth", is extremely extensive. An Internet search resulted in the following figures:

Term	results
Virtual City	1 million
Virtual Habitat	5 millions
Virtual Earth	81 millions
Digital Earth	135 millions
Virtual City	177 millions

Of scientific interest is of course just a small part. However the diversity of topics is tremendous. There are sensors and imaging techniques that have to be discussed, including various platforms like airplanes and drones, different types of geodata such as the cadastre or road maps, image organization, the determination of the exterior orientation, the 3D point clouds and the interpretation of facades and roofs widely studied in the field of computer vision. Within this thesis we have listed more than 350 references about these topics.

This chapter should summarize the state-of the-art in the field of 3D building models using LiDAR and aerial imagery and the semantic interpretation of these models by summarizing the most prominent publications for each field.

The first human documentation of geography can be traced back to about 15,000 years ago when people scratched the geography and surrounding on a piece of mammoth bone [Harley et al., 1987]. Since this time maps have been used as fairly convenient, efficient and effective representations of the environment or portions of it [Peuquet, 2001]. Among others, representing city environments and city objects has been a never-ending endeavor during the history of cartography.

Maps with more detailed outlines of building footprints served in public administration, primarily for ownership and taxation purposes. These maps (i.e. Tongguan map, [Liu, 1998]) can be seen as an early version of the cadastre. Later, for several centuries, cadastre diverted from forms of representation. Furthermore, more and more information about buildings beyond their footprint as contained in the cadastre, for instance, the properties of the buildings.

The revolution began when Waldo Tobler created the first "computer map" using a raster format [Tobler, 1959] in the late 1950s. 40 years later, after many breakthroughs in the computer technique, interest in modeling, managing and visualizing a city in a 3D environment had risen significantly, due to several reasons: (i) the development of computing technology; (ii) the availability of large, shared databases in digital form; (iii) the technique of high performance graphic cards; (iv) the advanced technologies for 3D data acquisition; (v) the efficient methodologies for generating 3D objects; (vi) the development of geospatial solutions and (vii) the miniaturization leading to smart phones and augmented reality.

Since the early beginnings, applications for spatial data are rapidly evolving beyond traditional cartographic products. In the areas of computer vision and graphics there is an emerging interest in the generation and visualization of synthetic environments that are accurately linked to the real physical world. This development is driven by the needs of defense agencies, the telecommunication industry and public institutions to get full 3D information [Gülch, 2000].

First researches in the field of aerial image understanding and interpretation started in the 1960s and 1970s [Binford, 1971; Nago et al., 1977; Nago et al., 1978; Nevatia et al., 1973; Nevatia et al., 1978] proposed different approaches to locate structures in aerial images. An early survey of image processing was done by [Huang et al., 1971].

In 1976, the Defense Advanced Research Projects Agency (DARPA) funded the Image Understanding Program (IU) with the goal to invest in research efforts to develop the technology required for automatic and semiautomatic interpretation and analysis of military photographs [Binford, 1976a; Binford, 1976b]. Therefore, every six months workshops were held to enable various program researchers to present interesting technical accomplishments achieved during the past six months. The main topics of this program were image processing, target classification and photo interpretation. One of the main objectives was to build computer systems that could locate and monitor buildings, airfields, aircraft and vehicles in aerial images [Binford et al., 1977]. Other works in the field of photo interpretation were presented by [Selfridge et al., 1981; McKeown et al., 1981; Lowe et al., 1981; Price, 1981]. At this time the centers of excellence

and their research leaders in the field of image interpretation that took part in the IU program were the Carnegie Mellon University (T. Kanade), University of Maryland (A. Rosenfeld), Massachusetts Institute of Technology (M. Brady), SRI International (M.A. Fischler), Stanford University (T.O. Binford), University of Southern California (R. Nevatia) and the University of Rochester (J.A. Feldman) all located in the United States.

Within this research many different programs were executed. One of the most prominent in the field of photogrammetry and computer vision was the RADIUS project as described in [1]:

"Research and Development for Image Understanding Systems (RADIUS) is a project of CIA Office of Research and Development, with support from the Advanced Research Projects Agency (DARPA). This highly applications-oriented project is intended to provide image understanding (IU) technology in a fully and semi-automated support system of human-machine interface interactive tools to the photo interpreter and imagery analyst (IA). It is based on the concept of a two- or three-dimensional "site model" that is used by image understanding algorithms for change detection, counting, and visualization. The project focuses on two aspects of Model-Supported Exploitation (MSE), (1) semi-automated and automated site model construction, and (2) exploitation based on site models. Model-supported exploitation (MSE) allows an image analyst to obtain a correspondence between an image and a stored three-dimensional model of a site. Automatic change detection and trend analysis can be performed based on image analyst needs. Three-dimensional site models provide a highly constrained context for applying IU techniques reliably.

The goal of the project was to develop a system that can be transitioned to the image analyst within 5 years. A Phase I study (Hughes, 1993-1994) developed the concept of operations and other requirements. Hughes Aircraft Company teamed with BDM International, Inc. (BDM), Washington, D.C., Control Data Corporation (CDC), Minneapolis, Minnesota, Hughes Research Laboratories (HRL), and the University of Southern California (USC). Using the results of Phase I, a two-year Phase II implementation project was initiated in late 1994 (Martin Marietta). An initial implementation was delivered to the NEL in late 1994, with updates delivered over the next two years."

For more details about the RADIUS project please refer to [Firschein et al., 1997; Nevatia et al., 1997a]. In the 1980s, research was focused on building detection and extraction from high resolution remote sensing imagery to extract semantic information from these images.

[Mohan et al., 1989] used perceptual organization to detect and describe buildings in aerial images. They detected linear features to generate rectangle hypotheses. Another method was proposed by [Huertas et al., 1988] that incorporates shadows to confirm hypotheses and to estimate the height of buildings. Other approaches that focused on building detection and extraction at this time were proposed by [Herman et al., 1986; Venkateswar et al., 1990; Irvin et al., 1989 Liow et al., 1990; Fua et al., 1987]. These approaches segmented the scene into regions to find edges lying on region boundaries and then looked for evidence of geometric structure among these edges to classify the region as a man-made object.

[1] http://www.fas.org/irp/program/process/radius.htm

Several photogrammetric conferences in the 1990s were devoted to the subject of automated feature extraction like the ISPRS Conference in Munich [Ebner et al., 1991], the ASPRS Conference in Washington D.C. [ASPRS 1994], the Ascona I Conference [Grün et al., 1995], the Ascona II Conference [Grün et al., 1997], and a 2nd ISPRS Conference in Munich [Ebner et al., 1999].

In 1998 [Brunn, 1998] classified building reconstruction into two main parts, namely, building detection and building reconstruction as cited in [Rottensteiner, 2001]. The building reconstruction strategies can be split into two approaches [Brunn, 1998]:

- Parametric – Top-Down Approach
- Generic – Bottom-Up Approach

Parametric Approach

This approach is used when the input data is not accurate enough to determine the shape of a building. Therefore, prior knowledge must be used to provide a robust reconstruction of the building. The topology of these primitives is provided by a database containing different primitive types. These model-based approaches are called top-down approaches as they try to fit constrained building models to the data. However, their complexity is restricted to the number and shape of primitives stored in the data base [Maas et al., 1999; Haala et al., 1999].

Generic Models

The feature based approach is used when input data is accurate enough for a robust reconstruction. Planar or curved segments are fitted into the raw data to create a polyhedral surface. These approaches are called bottom-up due to this constructive approach [Dorninger et al., 2008; Kada et al., 2009; Milde et al., 2009; Rottensteiner, 2010].

[Tarsha-Kurdi et al., 2007a; b] performed a comparison between data- and model- driven approaches for building reconstruction. It states that model-driven approaches are faster and don't deform the building model. Data-driven approaches tend to model each building detail to obtain the nearest polyhedral model, by deforming the building.

In most of the cases for automatic building reconstruction, model-based and feature-based approaches have to be combined to generate robust and accurate results. Depending on the used accuracy of the input data the appropriate approach has to be used. Most of the city modeling approaches directly or indirectly tackle the problem through point cloud analysis. Digital Surface Models (DSM), which are 2.5D view-dependent representations, constitute structured point clouds having a regular point distribution in the XY-plane well adapted to aerial-based large city modeling.

2.2. 3D Building models generated from imagery

In the 1990s research was focused on the automation of measuring geometric entries from digital images [Lin et al., 1994; Roux et al., 1994; Berthod et al., 1995; Gruber et al., 1995; Hsieh et al. 1996a; Hsieh et al., 1996b; Grün et al., 1998; Huertas et al., 1998; Gülch et al., 1999; Lammi, 1999; Li et al., 1999; Vosselman et al., 1999]. Other research was focused on

building roofs. The most promising developments in this field in the 1990s came from [Jaynes et al., 1994; Baillard et al., 1999b; Marengoni et al., 1999; Moons et al., 1998; Fisher et al., 1998; Henricsson, 1998]. Often, stereo images taken from aerial or satellite photography are used to generate a DSM. In former times, the produced DSM was in most of the cases of lower quality than from LiDAR data [Brenner, 2000]. One of the outstanding works in the field of cartographic feature extraction at this time was proposed by [Quam et al., 1991]. All these works were focused on the semi-automatic modeling and manipulation of 3D building models using imagery. [Nevatia et al., 1997b] described a method to detect and reconstruct rectilinear buildings with flat roofs from single aerial images. [Baillard et al., 1999a] proposed a method to reconstruct buildings in 3D from multiple images. Therefore, line features were extracted and matched by a method introduced by [Schmid et al., 1997]. These matched lines finally lead to piecewise planar reconstructions of the buildings. [Ameri, 2000] presented another approach that segmented Regions of Interest (ROI) for individual buildings by the method proposed in [Brenner, 2000]. Other methods were introduced by [Elaksher, 2002; Samadzadegan et al., 2005]. [Taillandier et al., 2004] that exploited cadastral maps, aerial images and DSM for generic modeling. Recent works from [Lafarge et al., 2010] and [Zebedin et al., 2008] generated DSMs from MultiView Stereo (MVS) imagery in order to model buildings using polyhedral structures. [Zebedin et al., 2008] presented a concept based on fusion of feature and area information for building modeling that relies on extracting geometric prototypes, such as planes, taking into account range data, 3D lines and an individual building mask. [Kluckner, 2011] proposed an approach that is mainly based on the availability of a semantic interpretation, where building regions can be easily extracted.

2.3. 3D Building models generated from LIDAR data

Ever since airborne LIDAR data appeared as a new data source in remote sensing and photogrammetry, researchers addressed the problem to model and characterize buildings using this data source. One of the main reasons was its performance for the highly automated acquisition of digital surface models (DSM) as well as the generation of digital terrain models (DTM). Its results for DTM generation in wooded areas were one of the main advantages of LiDAR [Kilian et al., 1996; Kraus et al., 1998]. The first approach that used only LiDAR data for building reconstruction was introduced by [Förstner et al., 1995]. Their method used two different 3D models: a parametric model for buildings with a rectangular shape and a prismatic model for complex buildings. Another approach was presented by [Maas et al., 1999] where building models were generated based on the analysis of invariant moments of segmented regions. Therefore, they determined building parameters like position, orientation, length, width and roof inclination using the first and second order invariant moments. [Stilla, 1999] presented a building reconstruction method based on the integration of maps for the extraction of the building outlines. Their approach generated a prismatic model by using the height values inside the building outline. The building roofs were classified into 4 categories: flat roofs, flat roofs with superstructures, gable roofs and hip roofs. [Vosselman, 1999] introduced a method that uses polyhedral building models and plane adjustment to derive a complete building description. Another approach proposed by [Alharthy et al., 2002] generates prismatic buildings based on edge detection. The approaches from [Geibel et al., 2000] and [Rottensteiner et al., 2002] are segmentation based that aim to find planar regions to construct a polyhedral building model.

[Gorte, 2002] proposed a segmentation based approach using a TIN structure. Segments are generated by iteratively merging triangles based on similarity measures.

The problem of most of these methods is that in dense urban areas the footprints have to be extracted manually. To overcome this problem existing map information has to be used. [Haala, 1999; Haala et al., 1999] proposed parametric building models based on predefined 2D building footprints. [Suveg et al., 2004] extracted the outlines of buildings using LiDAR, using a Hough transform. [Rottensteiner, 2007] introduced a method to determine building parameters. Therefore, in the adjustment of the model geometric regularities were included as soft constraints. He uses robust estimation to eliminate false hypotheses about these regularities.

Recent approaches in 3D city model generation using unstructured point clouds were presented by [Matei et al., 2008; Poullis et al., 2009; Toshev et al., 2010; Zhou et al., 2010]. [Matei et al., 2008] proposed a large scale building segmentation for densely classified urban scenes. Their approach preserves object boundaries and enables an accurate modeling of the rooftops. [Poullis et al., 2009] employed a three step approach, including preprocessing, segmentation, and modeling for a large scale reconstruction of 3D city models.

[Lafarge et al., 2006] proposed a method that started with simple primitives, based on rectangles, which are placed based on a digital surface model, using a Markov chain Monte Carlo simulation. In the following, the rectangles are merged to generate ground plans, which are then additionally subdivided along height jumps. From this, the final roof reconstruction is obtained. [Lafarge et al., 2006] introduced a method to generate rectangular building footprints in the surface models that uses symmetry criteria to roughly estimate the geometry of rooftops. In [Lafarge et al., 2010] extended this approach with a library of 3D blocks for improved building generalization from single DSMs. These blocks can be seen as single parts stacked together for building construction and have to be given in advance. A coarse estimation of the rooftop landscape can then be used to detect structures at a finer scale [Dornaika et al., 2008]. In 2011, [Lafarge et al., 2011] proposed a method to simultaneously reconstruct buildings, trees and topologically complex grounds.

Many other recent works in this field are devoted to street level modeling from ground or oblique view data, like [Banno et al., 2008; Frueh et al., 2004b; Furukawa et al., 2009; Strecha et al., 2008] or to point clouds generated from Internet-based multi-view stereo [Frahm et al., 2010; Agarwal et al., 2009].

2.4. Building Models and 3D City Models

A 3D City Model is a database of 3D urban objects that consists of buildings, vegetation, ground surface and water that produce a digital representation of a city. These models are digital, geo-referenced representations of objects and structures. Recently, these models are extended by semantic concepts such as CityGML. In contrast to models that are solely created for visualization purposes, semantic models store additional information about type, usage, and role of objects as defined by an predefined underlying ontology [Kolbe, 2009]. But also other semantic information can be stored in such models like the characterization of facades, the characterization of the building roofs and the characterization of the surrounding vegetation.

These models can be used not only for visualization purposes but also for complex analysis or thematic or spatial queries. Depending on the accuracy of the input data and the application, their level of detail (LOD) may vary widely. Buildings can be modeled as parallelepipeds with metric accuracy, or modeled in detail with all the 3D facets of their roofs and facades or even their interiors, with a centimetric accuracy. This variety of details that can be part of a 3D city model opens a large field of applications. [Fan, 2010] has listed in his work the following applications:

"Urban planning

Many of the existing models serve the main purpose of supporting urban planning processes by means of visualization of virtual scenes. Often, administrative departments find themselves confronted with complex decision processes of large-scale reconstruction projects of old town areas and investment projects (e.g. a new shopping mall, commercial area, industrial site). Using a 3D city model, the current situation of the city or the involved city part can be visualized as an overview. Besides 3D building models and a terrain model, various 3D city facilities like transportation objects, vegetation objects etc. can also be described in this model whose interactive visualization helps to present and evaluate the visual impact of planned constructions.

Civil engineering calculations

Civil engineering calculations require 3D building models with the highest possible details, preferably the architectural models with detailed wall and roof structures, balconies, bays, outlook as well as the internal organization. These detailed models allow calculations of volume and area of part of the building for cost estimation.

Safety and security

Detailed 3D city models containing buildings and other city objects can facilitate accurate positioning and inspection of incidents, which may relieve the lifesaving rescue work in emergency cases and result in a reasonable situation of awareness and safety among the inhabitants.

Disaster management

3D city models with detailed digital terrain model (DTM) in combination with computational fluid dynamics can serve as a basis for numerical flood simulations to predict whether parts of a city will be affected by a flood and how severe the impact on the buildings will be. These simulations help to plan efficient flood protection measures. Moreover, flooding can be simulated more precisely if there is sufficient thematic information embedded in the 3D city model.

Noise modeling and investigations

It is well known that vertical topographic structures, both artificial and natural, may affect noise propagation in a city area. A 3D city model is useful for the simulation of noise propagation over a wide area and investigation of the population under the harmful influence of noise. Contaminants in the air or line of sight for planning of radio network etc. can also be calculated.

Route planning and navigation system

Cognitive psychology studies have shown that persons who give directions use more than just basic elements such as street names and curves, but also landmarks such as orientation points. On this note, suitably textured 3D city models and suitably enhanced 3D landmarks are necessary to ensure the visual recognition of the urban environments for users of route planning and navigation systems. Moreover, interior structures of public buildings, park houses, and subway stations can also be modeled for indoor navigation.

Business development and tourism

3D visualization of terrain and city models including buildings, streets and other city furniture combined with tourism-relevant data (e.g. hotel and restaurant information, shops etc.) can support the potential visitors to make optimal decisions and increases the occupancy within the tourist region."

More than two decades of research have driven down the costs of building such databases. The attainable LOD of the resulting 3D city model depends on three parameters: the richness of the input data and the time and cost allocated. To ease the use of these 3D models, the City Geography Markup Language (CityGML) was introduced [Kolbe et al., 2005]. It is an XML-based data format that categorizes building models depending on their LOD in five different LODs (see Figure 3). It enables the exchange of spatial data including attributes, relations and geometry. Since this time, CityGML has established itself to a common standard for 3D city models. For more details please refer to [Kolbe et al., 2005]. Table 2 illustrates the descriptions for the different LODs for these 3D models.

	Description	*Accuracy (location)*	*Accuracy (Height)*	*Minimal sizes*
LOD 0	2.5 DSM	-	-	-
LOD 1	Block models without roof structures and textures	>5m	>5m	6m * 6m
LOD 2	Models with roof structures and textures	2m	1m	4m * 4m
LOD 3	Detailed models of facades and roofs (balconies, bays, overhangs, etc) with high resolution texture	0.5m	0.5m	2m * 2m
LOD 4	Detailed models of building interiors (rooms, doors, stairs, furniture, etc.)	<0.2m	<0.2m	-

Table 2: Description of different levels of detail including their accuracies [Kolbe, 2006]

To be able to satisfy the needs of an application, a generalization of the input data is necessary. This is not only important for creation of 3D building models, but also for their characterization. The automatic generalization of building models has been a research topic ever since. [Staufenbiel, 1973] proposed a method for the iterative simplification of 2D ground plans. Several algorithms have been developed to remove smaller line segments under a pre-defined length by extending and crossing their neighbor segments and by introducing constraints about

their angles and minimum distances [E.G.POWITZ, 1993; REGNAULD ET AL., 1999; VAN KREVELD ET AL., 2001; HARRIE, 1999; WEIBEL ET AL., 1996]. Other approaches use vector templates [MEYER, 1989], morphological operators like opening and closing [LI, 1996], least-squares adjustment [SESTER, 2000] or techniques from scale space theory [MAYER, 1998].

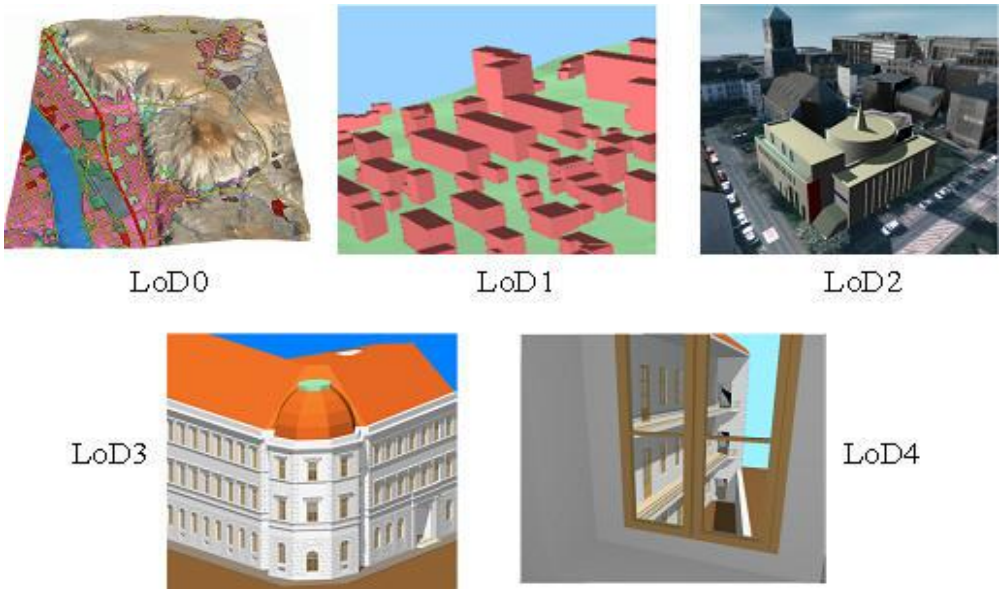


Figure 3: 3D city models generated using different LODs [KOLBE, 2006]

Nowadays, a few algorithms also exist that have been specifically designed for the generalization of 3D building models. [FORBERG, 2004] introduced an algorithm that adapts the morphology and curvature space operators of the scale space approach to work on 3D building models. Another approach was proposed by [ANDERS, 2005] that works for linearly arranged building groups. Their 2D silhouettes, which are the results of three projections from orthogonal directions, are simplified, extruded and then intersected to form the generalized 3D model. [THIEMANN ET AL., 2006] presents adaptive 3D templates. Building models are categorized into a number of classes with characteristic shapes. A building model is then replaced by the best fitting 3D template. Because the semantics of the template is known, the object itself or specific features of the model can be emphasized at will.

The simplification of 3D models has been a major topic in the field of computer graphics. See e.g. the survey of [LUEBKE ET AL., 2002] for an up-to-date summary of the most important work in this field. The progress 3D city modeling and the resulting increased availability and accuracy of 3D city models can be attributed to recent developments in remote sensing technologies and data extraction algorithms. Nowadays, it is possible to automate the reconstruction of 3D city objects [BECKER ET AL., 2008; HAALA ET AL., 1999; ROTTENSTEINER, 2007] and virtual 3D city models to a large degree. Thus the cost for the creation of large and city wide 3D city models have dropped rapidly during the last years. Consequently, many cities have started to build up 3D city models as part of their local data infrastructure. These models represent not only a virtual 3D city model, but are linked to the land cadastre system and thereby represent the city as it is described by official and legally binding geo-information.

2.5. Building Facades

Most approaches generate 3D city models with distinctive roof structures but with flat facades. However, in addition to the available building footprint and roof shape, facades are one of the most descriptive features that reflect the building style. [Lorenz et al., 2006] introduced an approach to add texture for each surface polygon. Depending on the distance, the viewing angle and occlusions the best image is determined for coloring the facade pixels. [Haala et al., 2010] stated that oblique aerial images are necessary to guarantee the required visibility. Within the scope of this work we will demonstrate that vertical aerial images are sufficient for this task, and we show that oblique images have many drawbacks due to occlusions generated by their viewing angles.

Since structures like windows, doors, balconies, bays etc. are also required, such a system has to identify them. Most building facade analysis techniques try to extract specific shapes/objects from the facade: windows frame, etc. Most of these methods are data driven, i.e. image features are first extracted and then some models are matched with them to build object hypotheses [Haider et al., 2010; Haugeard et al., 2009; Lee et al., 2004]. Other model-driven techniques try to detect more complex objects that are patterns or layouts of simple objects [Korah et al., 2007; Reznik et al., 2007; Han et al., 2005]. Higher level techniques use grammar based approaches to generate directly a hierarchy of complex objects composed of patterns of simple objects. Those methods generally devote their strategy to a special architectural style [Alegre et al., 2004; Müller et al., 2007; Ripperda et al., 2008].

[Lee et al., 2004] extracts window candidates defined as homogeneous rectangles from ground view images. They use an explicit window model and horizontal and vertical profiling of these facade images to find repetitive patterns. [Mayer et al., 2005] presented another image based approach aiming at the characterization of building facades. Another approach was introduced by [Haider et al., 2010] that extracts windows using an Adaboost algorithm. Both methods use an implicit window model and a learning approach to detect windows. In a similar fashion [Wenzel et al., 2008] uses a clustering procedure based on appearance similarity.

[Cech et al., 2007] proposed a segmentation based on a maximum posteriori labeling. Thus, each pixel is associated with values linked with some configuration rules and a set of non-overlapping window panes hypotheses is extracted. In contrast, model driven facade characterization methods use shape grammars [Stiny et al., 1972]. These methods rely on the similarity and repetitive patterns of facades. [Wonka et al., 2003] introduced an approach that allows the automatic generation of architectural structures from a database of rules. Shape grammar generation is the basis for [Müller et al., 2007] work that aims to generate large scale city models. [Müller et al., 2007] uses intensity based similarity measures including cross correlation and mutual information. The basis for the approaches presented by [Müller et al., 2007;Wahl et al., 2007; Russakoff et al., 2004;Pluim et al., 2000; Studholme et al., 1999] is that windows are rectangular structures. Methods exploiting boundary information can also be divided according to their inner architecture. [Ripperda et al., 2008] fix the grammar rules according to prior knowledge: they compute the distribution of facade elements from a set of facade images. Other works in the field of facade reconstruction are presented in [Hohmann et al., 2009; Burochin et al., 2010; Becker, 2009].

Single layered methods using entities acquired directly from an image. [Opelt et al., 2006] uses boundary fragments with information about object's centroid as weak detectors in a modified AdaBoost scheme. [Leordeanu et al., 2007] model pairwise spatial relations between boundary fragments in a robust redundant way and use this compositions in a logistic classifier considering their relevance and match quality. [Gu et al., 2009] propose a method that uses object regions for determining the hypothetical bounding box of the object. [Donoser et al., 2010] proposed a method for getting stable object regions suitable for object recognition using Hough voting scheme. Multiple-layered hierarchical methods work with hierarchically organized shape information. [Ullmann et al., 2007] proposed to acquire highly informative fragments of an object and split them recursively down to atomic parts. An abstract object model is generated by exploiting the similarities and co-occurrences of the parts, A method proposed by [Ommer et al., 2010] models objects using a hierarchy of compositions of SIFT-like descriptors and Bayesian networks.

In the LiDAR field there is a vast amount of work done on range data segmentation [Hoover et al., 1996; Jiang et al., 1996]. [Wang et al., 2011] states that: "*However less work has been done on detecting windows from LiDAR data. This is probably because of the nature of LiDAR data: noisy and sparse. Microsystems like windows are hard to differentiate. In addition, windows without curtains often retain signals from the interior of the buildings and the returned signal does not always contain enough valid data.*" [Wang et al., 2011] also states that there are two major works on window detection using LiDAR data [Pu et al., 2007; Haider et al., 2010].

2.6. 3D Reconstruction

Image based 3D reconstruction of a scene from multiple overlapping views is a fundamental and well studied problem in computer vision [Hartley et al., 2004] and photogrammetry [Kraus, 2004]. The photogrammetric reconstruction problem can be roughly divided into three processing steps:

Structure from Motion and Aerial Triangulation

The problem of structure from motion computation, i.e. recovering the (sparse) structure of a 3D scene and camera orientations, has reached maturity over the past years. In the photogrammetric literature, structure from motion (SfM) is denoted as aerial triangulation (AT) and is a well known part of the traditional photogrammetric processing chain. It consists of the manual or automated selection of point measurements in overlapping images (tie points and ground control points) and the determination of unknown parameters of the camera pose and position by means of photogrammetric bundle adjustment [Brown, 1976]. Today, prototypes of fully automatic processing chains for ordered [Polleyfeys et al., 2004] and unordered image collections exist [Martinec et al., 2007; Irschara et al., 2007]. Nowadays, it is even possible to automatically reconstruct a scene from unorganized and very inhomogeneous image datasets such as community photo collections gathered from the web [Snavely et al., 2006]. The success of these methods is mainly based on substantial progress in wide baseline matching and local scale invariant feature detectors [Lowe et al., 2004], large scale image retrieval [Nister et al., 2006], advances in multiple view geometry [Triggs et al., 1999; Nister, 2004], and the leverage of cluster computing power [Agarwal et al., 2009]. In aerial photogrammetry the wide baseline

matching problem is not explicitly addressed, since approximate knowledge of the exterior camera orientations is normally provided by global positioning systems (GPS) and inertial measurement units (IMU). Thus, in the photogrammetric workflow, AT is mainly a refinement step, necessary to get sub-pixel accurate camera orientations [Leberl et al., 2010b].

Dense Matching

Structure from motion yields camera orientations and a sparse set of point features, but for practical applications like DSM extraction and 3D modeling a depth estimate for every individual pixel is required. Today's photogrammetric flight missions are usually based on high image overlap [Leberl et al., 2003b]. This means that on average a point in the scene is visible in ten to twenty images (depending on the degree of occlusion). Consequently, the redundancy implies that multi-image matching approaches [Okutomi et al., 1993;Stuehmer et al., 2010] can be employed that directly leads to a reduction of problems caused by occlusions, surface discontinuities and results in higher measurement accuracy through the intersection of more than two rays.

There exists a variety of algorithms and methods addressing the dense matching problem in computer vision. A good survey of existing stereo matching methods is given in [Scharstein et al., 2003] and [Bleyer et al., 2010] where they extensively investigate the relevance of color information for matching. Depending on the set of pixels where the optimization of the disparity values is performed, dense matching methods can be roughly categorized into three classes. Among them are local methods [Birchfeld et al., 1998; Yoon et al., 2006], semi-global methods [Hirschmüller, 2008] and global optimization methods [Pock et al., 2008].

Many dense stereo matching algorithms are based on plane-sweep techniques [Cox et al., 1996; Gallup et al., 2007]. The plane sweep concept allows the direct accumulation of matching costs [Hirschmüller et al., 2009] for multiple images. In Figure 4, a result of such multi-view plane sweep method based on global disparity optimization [Pock et al., 2008] is shown. This method delivers global optimal solutions and is easily parallelizable for the execution on current GPUs; however, this method requires a large amount of memory. Other stereo matching approaches rely on image rectification techniques [Fusiello et al., 2000]. For a rectified stereo pair, the stereo correspondence search can be restricted along a horizontal line [Hartley et al., 2004], thus stereo matching reduces to a 1D optical flow problem [Zach et al., 2007a]. Two-frame stereo matching algorithms are well studied methods [Scharstein et al., 2003]. Recently, [Strecha et al., 2008] has released a similar benchmark for multi-view stereo evaluation. The dataset consists of high resolution images, and a provided ground truth enables a detailed and accurate evaluation of multi-view matching methods. Robust (optimal) multi-view matching that fully exploits the available image information is still an open research problem.

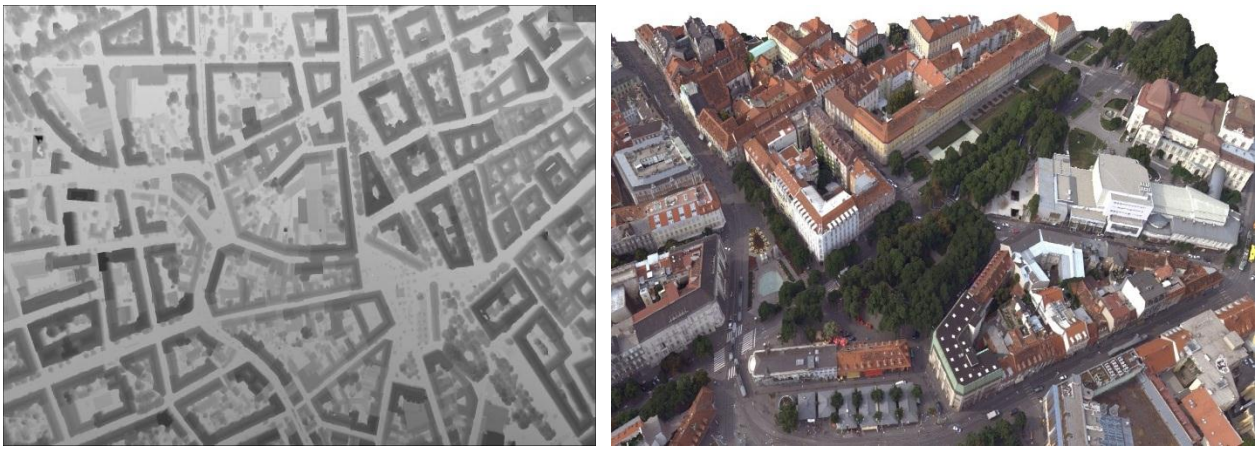


Figure 4: Range image (depth map) of an airborne image of Graz and oblique view of a texturized depth map. Depth map and textured 3D model were created using Institutes Dense Matching Pipeline.

Range Image Fusion

Dense matching techniques produce 2.5D range images, also denoted as depth maps that represent the distance for each pixel to the estimated surface of a scene. The inherent redundancy of digital aerial images allows combining and fusing several range measurements to one and the same point of a 3D scene. Depending on the target application, range image integration can be either done in 2.5D [Zebedin, 2010] or in full 3D space [Curless et al., 1996; Zach et al., 2007b]. Furthermore, the application of range image integration is not only limited to the fusion of stereo depth maps, but can also be used for processing range data from active sensors like LiDAR (Light Detection and Ranging) [Hilton et al., 1996] and time of flight cameras [Ullrich et al., 2002].

Robust range image integration is often based on an intermediate volumetric representation that allows the generation of models with arbitrary genus and avoids the numerical difficulties encountered with polygonal techniques [e.g. Hiep et al., 2009]. Early work on volumetric range image integration employs an averaging scheme of 3D distance fields to combine several 2.5D range images [Curless et al., 1996; Hilton et al., 1996]. Hence, the obtained surface is basically the minimizer of an underlying energy function with quadratic data fidelity terms. As pointed out in [Hornung et al., 2006], simple averaging without further regularization causes inconsistent surfaces due to frequent sign changes of the mean distance field. [Zach et al., 2007b] address this problem by using an energy functional consisting of a total variation (TV) regularization force and a L1 data fidelity term. It was also shown in [Zach, 2008] that such an approach can be greatly accelerated by taking advantage of graphic processing units. [Merreli et al., 1981] even demonstrated real-time performance for range image integration by leveraging the massive processing power of GPUs.

Variational Methods for 3D reconstruction

3D reconstruction aims at modeling the inverse process of image formation. However, in a mathematical sense, the inverse projection mapping does not exist, since all 3D points along a

visual ray are projected onto the same image point [Kolev et al., 2009]. This makes the problem of multi-view 3D reconstruction ill-posed, and additional regularization constraints are needed in order to get smooth surfaces. Variational methods for multi-view 3D reconstruction pose the problem as one of modeling the continuous two-dimensional surfaces in space by minimizing an appropriate energy functional [Keriven et al., 1998; Jin et al., 2005]. The incorporation of a total variation (TV) regularization term allows the extraction of minimal 3D surfaces and thus regularizes the resulting 3D model in a proper way [Chan et al., 2004].

Recently, [Pock et al., 2008] introduced a convex formulation of continuous multi-label problems which can be applied to stereo matching and delivers globally optimal solutions. In contrast to graph-based methods [Boykov et al., 2001; Kolmogorov et al., 2004] that can be used to solve the same problem by minimizing a discrete energy, the continuous formulation assures that the algorithm does not suffer from metrication errors. Furthermore, the continuous formulation is memory efficient and the underlying numerical optimization scheme can be easily parallelized on current GPUs, therefore ideally fitting to 3D facade reconstruction from aerial images.

2.7. Real property evaluation using aerial images

In the course of a comprehensive investigation it was not possible to find any publications in connection with automated property valuation using aerial images. The only publications were about the automatic valuation of forests from aerial images [Avery et al., 2001; Bullard et al., 2000; Gunter et al., 1996; Helms, 1998; Husch et al., 2002; Paine et al., 2003; Smith et al., 1996].

CHAPTER THREE

FROM CHARACTERIZATION TO VALUATION

Contents

3.1	Some general definitions for Real Property Valuations	53
3.2	What can aerial photos contribute?	56
3.3	From the characterization to a valuation	57
3.4	Proposed Processing Framework	58
3.5	Experimental Datasets	61

3.1. Some general definitions for Real Property Valuations

Galileo Galilei once said[2]: "*Measure what is measurable, and make measurable what is not so*."

We want to measure "value" of a real property, certainly a task belonging to the second part of Galileo's statement, namely making "measurable" what is normally not obviously measurable. Whenever real estate is bought or sold, a mortgage loan is given, or a property will be developed, the valuation of the property is a central factor in planning and decision making, as well as in all financial aspects of a project.

To ease the characterization of real properties, let us review the meaning of the word "Property" and "Building" using the Austrian juridical framework (Vermessungsgesetz):

„*Ein Grundstück ist in Österreich jener Teil einer Katastralgemeinde, der im Grenzkataster als solcher mit einer eigenen Nummer (Grundstücksnummer) bezeichnet ist. Grundstücke werden durch Beschluss des Bezirksgerichts, welches das Grundbuch führt, neu gebildet oder gelöscht. Sie können auch durch Teilung, Zusammenlegung, landwirtschaftliche Regulierungsverfahren (Kommassierung) entstehen. Grundstücke sind Teile eines Grundstückskörpers (Liegenschaft) und können aus Parzellen (geschlossenen Einheiten ein und derselben Nutzungsart, bezeichnet durch die Katasternummer) bestehen.*"[3]

[2] www.thinkexist.com
[3] www.bev.gv.at

A property consists of several entities:

- The Land,
- The Natural Growth (§§ 295, 405 ABGB, thus the vegetation),
- The buildings constructed on that land,
- The components which are connected with the property (ground-fixed, pipes, wires etc.),
- Attachment (the movable property which should remain with the property)

In our case we are especially interested in the land and the buildings that are constructed on that land. A building is defined according to the Bundesgesetz "über das Gebäude- und Wohnungsregister" (GWR-Gesetz), Article 1 as follows:[4]
"Residential buildings of all kinds (single-family houses, dwelling house arrangements, row houses, etc.), hotels and other buildings for short-term accommodation, office buildings, wholesale and retail trade buildings, buildings for transportation and communication (e.g., railway stations, terminal buildings, parking garage buildings, etc.) industrial buildings and warehouses (e.g., factory buildings, silos, containers, etc.), buildings for cultural and leisure purposes (e.g., museums, libraries, gymnasiums, etc.) as well as buildings for the educational system and health service industry (e.g., schools, research facilities, hospitals, etc.)."

The purpose of the valuation is to achieve a balance of interest between the "supplier" of real estate and its "consumer". Besides, frequently both parties operate with different priorities and value propositions. However, all relevant quantity and quality aspects of a property are expressed by one single number apart from exceptions like fiscal or actuarial values, the result reflects as a rule of the attainable market value.

There are also other values in connection with a property. Table 3 lists these values with their relevance in the daily life. As shown in Table 3, the market value is the value which is needed most often. By the closing of property insurances, banks determine the market value and deduce the hypothecary value. The real property market is affected by many parameters: recognizable or hidden, objective or subjective, inaccurate or correct. Therefore, it is often impossible to define the only proper "value". Tax assessment maybe algorithmically modelled with parameters from a characterization. For the determination of the market value of a real property there are three standardized valuation methods:

1. Comparative value method (Vergleichswertverfahren)
2. Gross rental method (Ertragswertverfahren)
3. Asset value method (Sachwertverfahren)

We extract a review of these three approaches from the webpage Wikipedia.[5]

[4] www.statistik.at
[5] www.de.wikipedia.org

Type of value	Relevance	Is determined by
Market value	Judgments Assessment of corporate shares Case of succession, divorce Purchase and sales negotiations Disproof of the values ascertained by the tax Dispossessions Compensations	Appraiser Advisory commitee
Hypothecary value	Hypothecation of real property	Banks
Insurance value	Closing of a building insurances	Insurances
Assessed value	Taxation of real property	Tax office

Table 3: Different values in connection to real property and their relevance [METZGER, 2006]

Comparative value method:

"With the comparative value method, the comparative price is derived as an average from purchase prices of comparable properties. Normally the comparative value method is not suited for the determination of the market value of a developed real property, because mostly insufficient current comparable transactions are available. For the registration of suitable transactions of not built-up real properties, it is necessary to first look through the property transactions in the title register for the appropriate purchase processes and then make a selection of relevant entries.

The following points are to be followed:

a) *A reliable judgment can be possible only with an adequate number of suitable purchase prices (at least 20 – 30 sale prices of real properties).*

b) *Comparative prices from the closer or further surroundings are often not available. Thus, reference prices from other areas can be used on condition that the local circumstances and the respective market situation is comparable (it is often not achieved).*

c) *The following value affecting factors are of interest for the comparability and for the adjustment of the purchase price listed in Table 1.*

However, we will argue that "value is reasonably available per square meter of land and of living space in a certain area."

Gross rental method:

"The gross rental method is applied if the market participants estimate the property to be valued under profit points of view. In this case, the value of the property is so determined that the basis of the value calculations are the prospected net earnings of the property. Depending on whether the property is built-up, the value of the whole property (ground plus building or only the value of the ground) has to be determined. However, in both cases you receive it from the net earnings of the property. In one case it has to be estimated on the basis of the existing building, in the other case on the basis of a fictive building."

Asset value method:

"In the case of a built-up property and the presumption that the construction (building) represents an essential value, the asset value method is an alternative possibility to determine the value of a property. An initial value will be set first, consisting of:

- *Land value*
- *Value of the physical structures*
- *Value of the outdoor facilities*

All three values are determined separately and are combined at the end. Taking into account value decreasing and value rising factors one finally receives the asset value of the property."

3.2. What can aerial photos contribute?

A rather fundamental question exists: Which external parameters of a property can be extracted from digital aerial images? Table 4 lists all external parameters which have a role in the evaluation process that can be extracted from vertical aerial images.

Computer vision is the approach or toolbox to unlock this information that is encoded in aerial photographs.

Computer vision *is the science and technology of machines that see. As a scientific discipline, computer vision is concerned with the theory for building artificial systems that obtain information from images…* [6]

Without computer vision the characterization of the visual contents of a picture cannot be automated. We argue that computer vision as it exists today offers a sufficient range of tools to address the real estate valuation problem. Therefore, the focus will have to be on testing existing algorithms for their applicability to property valuation.

[6] Wikipedia (12/2008)

Value affecting factors		
Parameter	Obtained from aerial images	Comments
Position to recreation areas, cultural or economic centers, industrial zones	Yes	The surroundings of a building can be determined using the image classification derived from the aerial images and/ or external databases (not discussed within this project)
Surface arrangement	Yes	Analysis of image classification for property (discussed in chapter 5)
Plant cover	Yes	Analysis of vegetation derived from CIR aerial images (not discussed within this project)
Front width and depth	Yes	Building footprint and orientation of the footprint (discussed in chapter 6)
Shape	Yes	Building footprint and elevation data
Floor space	Yes	(discussed in chapter 5 and 6)
Roof extension	Yes	Roof analysis (discussed in chapter 8)
Windows, floors and architectural features	Yes	3D Façade reconstruction and characterization (discussed inchapter 7)
Superstructures	Yes	Roof analysis (discussed in chapter 9)
Revenue (rental fees, etc.)	No	
Land charges	No	

Table 4: Determination of value affecting factors derived from aerial images

3.3. From the characterization to a valuation

How can we make the transition from characterization to valuation? We have to link together the whole system by using the knowledge about value and the extracted characteristics of a building. The linkage between the characteristics and the value is similar to the American website zillow.com that uses public property tax records (see Figure 2). Real estate agents in the US always look for comparable estates for their appraisals. The basis for these appraisals are data points. These data points represent already realized real estate transactions that are stored in a database that is updated in regular time intervals. If someone is interested in a certain building, all building transactions within a certain radius are analyzed to determine the appropriate value. Therefore, buildings with similar interior and exterior characteristics, like floor area, parcel area, rooftop extensions, number of floors, number and size of windows, age of the

building etc. are taken into consideration. So far, most of these characteristics are determined manually by a visual inspection. Within this project, we simplify the characterization of the building exterior and thus reduce the complexity and duration of such a valuation.

In our case, we extract this information from vertical aerial images. However, to be able to realize this semantic interpretation of these images a processing framework is necessary.

3.4. Proposed Processing Framework

Low cost by automation and by a reliance on pre-existing sources has led us to define a general processing framework (see Figure 5). We start out by conflating (merging) geometric data from essentially two sources: the aerial imagery and the cadastral information. Figure 6 and Figure 7 cover an example for a 400 m x 400 m urban test area in the city of Graz (Austria). Conflation defines each property as a separate entity for further analysis, not unlike previous authors have proposed [Kienegger, 1992]. The cadastral data may also contain preliminary information about a 2D building footprint. Conflation is part of a pre-processing workflow with results in all geometric data being available per property and in a single geometric reference system.

We now proceed towards the use of the dense 3D point clouds associated with the aerial photography and extracted from it by means of a dense matcher applied to the triangulated aerial photographs [Klaus, 2007]. "Dense" refers to the fact that each image pixel gets assigned a terrain elevation value. The surface model is thus at a spacing equal with the Ground Sampling Distance (GSD) of the images. First is the extraction of data per building and per element of vegetation. This finds the areas occupied by a building as well as its height. For vegetation we need to find the type, its location, the height and the crown diameter. The building footprints get refined vis-à-vis the cadastral prediction using image segmentation and classification to define roof lines.

From the building one proceeds to the facades: building footprints become facade baselines. A decision is needed on the number of facades. We segment a single building shape so that the footprints of its parts all consist of simple rectangular shapes. This is the basis for addressing the facades. For each facade we must find its intersection with the ground -- its footprint. This is the basis for an extraction of the facade in 3D by intersecting it with the elevation data. We compute the corner points of each facade. These can then be projected into the block of overlapping aerial photography. We can search in all aerial images for the best representation of the facade details; typically this will be the image with the maximum number of pixels for a given facade. Since there are multiple facade images, we prepare for a multi-view process.

What follows is a search for rows and columns of windows in the overlapping (i.e."redundant") photographic imagery. First of all, this serves to establish the number of floors. Second, we are also interested in the window locations themselves, as well as in their size. And finally, we want to take a look at attics and basement windows.

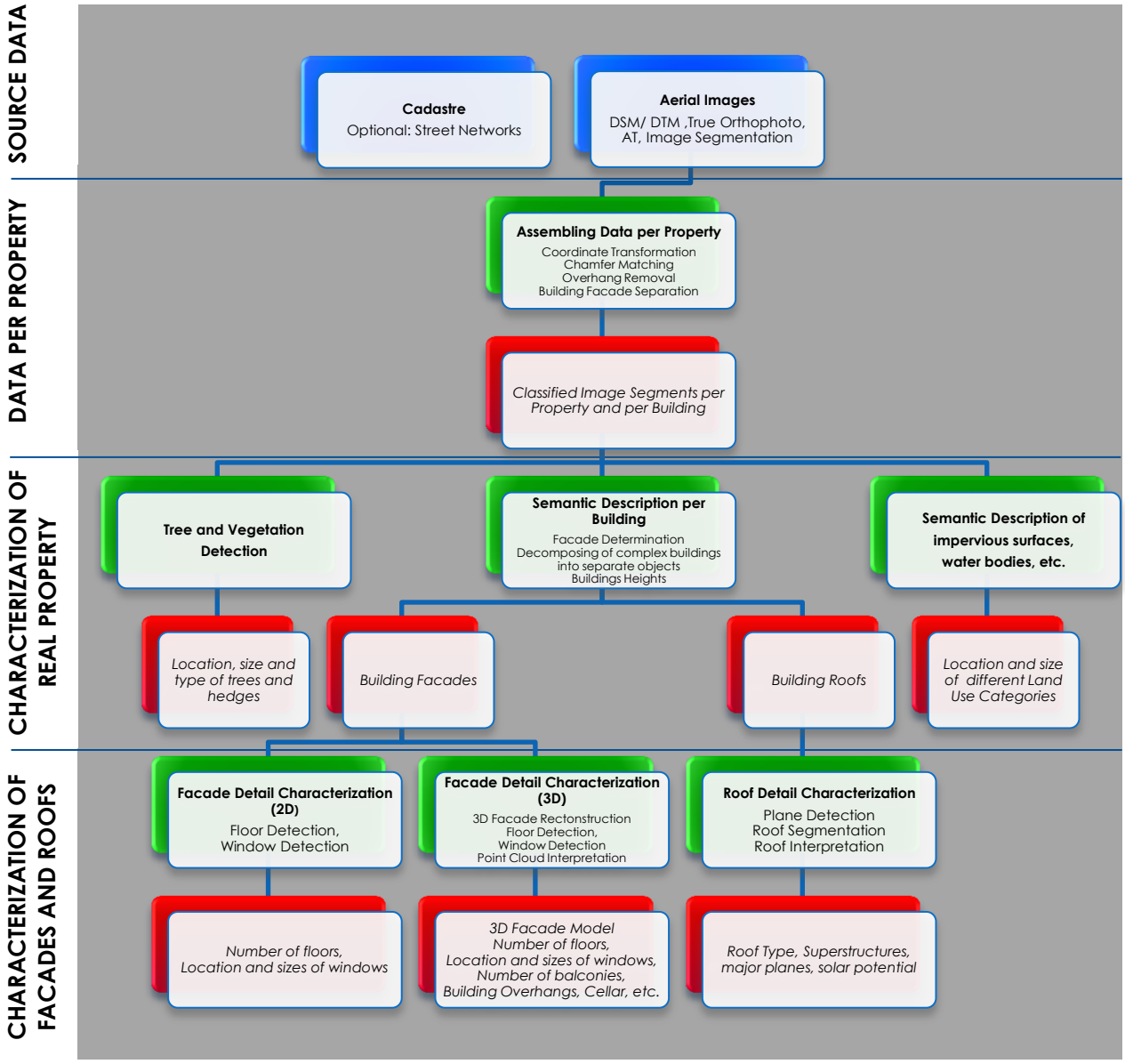


Figure 5: Diagram of the proposed work flow to characterize real properties from aerial images and associated cadastral data.

While Figure 5 summarizes the work flow, subsequent chapters of this work will fill in the details based on sample data from Figure 6, Figure 7, and from the city of Graz to explain the approach. The aerial coverage is taken with a ground sampling distance at 10 cm and with overlaps at 80% forward and 60% sideward. Our aim is to show that an automated property characterization is feasible, not least due to the fact that the overlaps produce a 10-image coverage of each terrain point.

Figure 6: True Orthophoto of the city of Graz, 400 m * 400 m, at a ground sampling distance of 10 cm. "True" in this orthophoto implies that the aerial photograph has been re-projected considering the elevation of the buildings so that one cannot see any of the building facades. Shown are 216 buildings Detail: Brockmanngasse, Kopernikusgasse; UltraCamX, acquired in 2007. (Courtesy Microsoft-Graz)

Figure 7: Cadastral map of Graz with legal property boundaries is shown for the area in Figure 3. The original data is in vector format that has been converted to a raster image at a GSD of 10cm. Shown are 300 separate parcels [Courtesy BEV-Austria].

3.5. Experimental Datasets

For our research we use two major datasets:

- Vertical Aerial Images and derived products like the DSM, DTM, True Orthophoto, Classification Image
- Cadastral Data

In the current application, we process aerial images taken by a large format digital aerial camera, specifically by the UltraCamX [Gruber et al., 2008b]. This, like most digital aerial cameras, produce images in four colors: red, green, blue and near infrared (NIR). These cameras also collect a separate panchromatic channel. The images often have ~ 13 bits of radiometric range; this is encoded into 16 bits per color channel, [Scholz et al., 2008; 2009].

For the evaluation of the proposed processing framework we use two major aerial imagery data sets and cadastral data:

- We use in this project the dense urban core and rural outlying areas of the city of Graz, Austria, covering 155km². This surface area is covered by 3000 aerial photographs that have been flown with overlaps that along-track are at 80% and across-track are at 60%. The Ground Sampling Distance (GSD) is at 10cm. Figure 8 shows a true orthophoto of the area covered by the aerial photographs.

Figure 8: True orthophoto of the city of Graz covering part of the dense urban core of the city [Kluckner, 2011].

- The second test area covers the dense urban area of the city of Annecy, France with irregular and complex rooftops. The area was covered using vertical aerial photographs with a GSD of 10cm. Figure 9 illustrates the downtown area of the city of Annecy.

Figure 9: Orthophoto showing the Downtown area of the city of Annecy, France flown with a GSD of 10cm [Laboratoire MATIS, IGN Paris]

- For the study area in the city of Graz we have cadastral data available from the year 2008, provided by the Austrian Federal Mapping Agency Bundesamt für Eich- und Vermessungswesen (BEV) (see Figure 10).

Figure 10: Cadastral Boundaries for part of the urban core of Graz. Green rectangles illustrate the four test areas.

In the next chapter, we will justify our choice of using vertical aerial images for the semantic characterization of urban areas. We will compare vertical aerial images with the other two main airborne data sources for 3D city model generation and its characterization and give a short description of the derived products that are necessary for that purpose. We will also discuss in more detail the role of the cadastral information in this context.

CHAPTER FOUR

SOURCES OF GEODATA

Contents

4.1	Introduction	65
4.2	Vertical Aerial Images	67
4.3	Satellite Imagery	72
4.4	Oblique Aerial Images	74
4.5	Aerial Laser Scanning (LiDAR)	76
4.6	Street Level Images and Lasers	78
4.7	Crowd Sourced Image Data & Community Photo Collections	79
4.8	Unmanned Aerial Vehicles (MAVs)	80
4.9	Catastre	81
4.10	Optional Street Network and Address Data	82
4.11	Derived Products	83
4.12	Experiments	87
4.14	Discussion and Summary	105

4.1. Introduction

One can observe an avalanche of digital geo-data coming from different kinds of sources that are used for the characterization and construction of 3D city models. A range of source data platforms and sensors exist: conventional airplanes, driving vehicles, remotely controlled aerial micro-vehicles (MAVs), or indoor tripod-based scanning and manual photography, GPS, IMU etc. It remains unclear which platforms, sensors and types of imagery are most efficient for the characterization and reconstruction of urban areas. For example, is imagery less efficient than point clouds obtained from LiDAR measurements taken from aerial platforms?

Because of the diversity of geo-data sources (see Table 5) in a first step we have to answer the following question: Can one analyze details of real properties from overlapping vertical or oblique aerial photography, airborne or terrestrial LiDAR, crowd soured imagery or Unmanned Aerial Vehicles (UAVs) and at what level of completeness and accuracy? One of the goals of our research is to clarify this question. However, an evaluation of all these different geo-data

sources would exceed the scope of this work; therefore, a detailed evaluation in the last part of this chapter focuses on airborne systems for semantic image interpretation.

The association with a geometric resolution or accuracy shows a range between ± 2 cm and ± 10 m and demonstrates the diversity of these data sources. Therefore, in the next part we want to discuss the different data sources in more detail and their role when dealing with 3D city models and their characterization.

OVERHEAD SOURCES	*URBAN GSD*
1. Satellite Imagery	0.5 m
2. Aerial Imagery, vertical	0.03 m
3. Aerial imagery, oblique	0.1 m
4. Aerial Laser Scanning (LIDAR)	0.1 m
STREET SIDE SOURCES	
5. Street Side Imagery from Industrial Systems	0.02 m
6. Street Side Lasers	0.02 m
7. Crowd-Sourced Images (FLICKR, Photosynth)	0.02 m
8. Location Traces from Cell Phones and GNSS/GPS	5 m
OTHER DOCUMENTS	
9. Cadastral Maps, ParcelMaps	0. 1 m
10. Street Maps from Car Navigation	5 m
11. Address codes with geographic coordinates	10 m

Table 5: The major sources of urban sources for geo-data and their expected typical geometric resolution. A considerable dynamic exists to conflate image, vector and address data, to create updating mechanisms for street-level and address data [Google blog on conflation and updating, [1, Blumenthal, 2009].

4.2. Vertical Aerial Images

Aerial Images that are produced in a nearly perpendicular direction are called vertical photographs or images. During aerial surveys nadir photographs will be taken from airplane-carried high precision cameras. There result, mostly regular, stripe-wise arranged picture formations in which adjacent images overlap. That means that an object is captured in at least two subsequent images. This provides the basis for the stereoscopic view and for further processing.

Vertically shot aerial images are well suited for photogrammetric evaluations, because the images map the terrain in a central perspective and the position and altitude of each image are approximately known. The model of the central perspective the relations that exist geometrically between the location of the points in the object space and their location in the image plane.

While highly overlapping color, infrared images and geometric data are assumed to be given in advance, additional products derived from these images are required to achieve the interpretation of the aerial scenes that will be discussed in more detail in a later section.

For the ease of reference in the next section, we give an overview about the development of vertical aerial images in more detail and how it became feasible to use vertical aerial images for semantic interpretation of building facades.

History of Vertical Aerial Images

The first known aerial photograph was taken by French photographer and balloonist, Gaspar Felix Tournachon, known as "Nadar", in 1858 from a balloon from an altitude of 80m over Paris. In these times, besides balloons, one could also use kites, pigeons and rockets to take cameras into the skies. After the earthquake in 1906 that destroyed large parts of San Francisco, a camera attached to a string of kites was used to document the devastation. The specially designed large-format camera provided panoramic images. These images remain some of the largest aerial exposures ever taken. During World War I, the need for aerial photography increased to produce battle maps and to serve as the "eyes of the army"[7]. "*England estimated that during the war its reconnaissance planes recorded around one-half million photographs. Germany calculated that if all of its aerial photographs were arranged side by side, they would cover their country six times*"[7]. During World War II, more sophisticated techniques in air photo interpretation were developed and strongly influenced the outcome of the war.

"In the 40s and 50s, the US, UK and Russian military undertook many surveys, including areas of mainland Europe. In the 60s, 70s and 80s, aerial photo surveys carried out on behalf of local and central government, utilities and large civil engineering projects became routine. During the 80s, the transition was made from black and white film to color film. Throughout this time many of

[7][http://northstargallery.com/aerialphotography/history%20aerial%20photography/history.htm]

these surveys were used on a 'one-off' basis, i.e. they were flown for one client who had one use for them. Resale of the photographs did happen, but it was still a relatively specialist subject".[8]

From Analog to Digital

"It was in the mid to late 90s that the resale value of aerial photography was really acknowledged. The shift was made from the survey companies carrying out a survey on behalf of a client, to the survey company investing in speculatively acquiring photography and making a commercial business from selling the photos multiple times"[8].

These aerial imagery improvements had many drawbacks compared with the upcoming LiDAR technology, The major problem of analog aerial imagery at this time was obvious: LiDAR was digital, aerial imagery was not! Additionally, analog aerial imagery was very expensive, difficult to process, and not very efficient. The aerial companies had to buy rolls of expensive stable (metric-quality) wide-format photographic film that had to be chemically processed after its exposure in the camera. The film processing itself was demanding and expensive: the use of lots of different chemicals having a short shelf life, a dedicated darkroom-cum-laboratory, and retainment of a highly skilled staff. After the development in a next step the analog images had to be scanned using an high-quality film scanner for further processing in an image processing system or a digital photogrammetric workstation (DPW) [Petrie, 2003]. Because of the high costs of film and the long exposure times, typical analog aerial images had a forward/side overlap of 60%/20%. Thus, resulting in insufficient redundancy was the major reason that automatically created DEMs using stereo matching had: i) significant errors and missing data and ii) occlusions due to a lack of redundancy, which was the major limitation of orthophoto products. The lack of redundancy made it almost impossible to characterize building details like facades because of very steep viewing angles and not enough redundant image data to solve this problem.

For these reasons the traditional applications for aerial images were limited to orthophoto production and aerial mapping. In contrast, LiDAR was thus used for DSM and DTM generation and the semantic interpretation and processing of these point clouds.

This changed when in 2004 the first large format digital aerial survey cameras were commercially used. The first digital aerial camera in this field came from Leica and was inspired from space cameras with a push-broom sensor (see Figure 11)." *It consists of an optical system projecting an image onto a linear array of sensors, typically a charge-coupled device (CCD) array. At any given time only those points are imaged that lie in the plane defined by the optical center and the line containing the sensor array"* [Zhang et al., 2005].

Another technology for digital cameras was introduced by Intergraph using a frame-based sensor. But to be able to compete with the traditional analog cameras, digital cameras had to offer an image width at least of 11500 pixels and a color band including RGB and infrared channels [Leberl et al., 2003b]. Thus, because of the limited size of a single charge-coupled device (CCD) array (34.7mm * 23.7mm for UltraCamX) a large frame format is achieved by combining several CCD arrays to satisfy this requirement. Another difference in frame systems is the color which is obtained at a lower geometric resolution than the panchromatic channel, and the final image is a result of fusing panchromatic with color bands. Forward motion

[8] [http://www.oldaerialphotos.com/History_of_Aerial_Photography.cfm]

compensation is implemented in some camera types to achieve longer exposure times and that results in better radiometry. Current large-format aerial cameras include multiple area CCD's and up to eight lenses [Leberl et al., 2003a; Gruber, 2007].

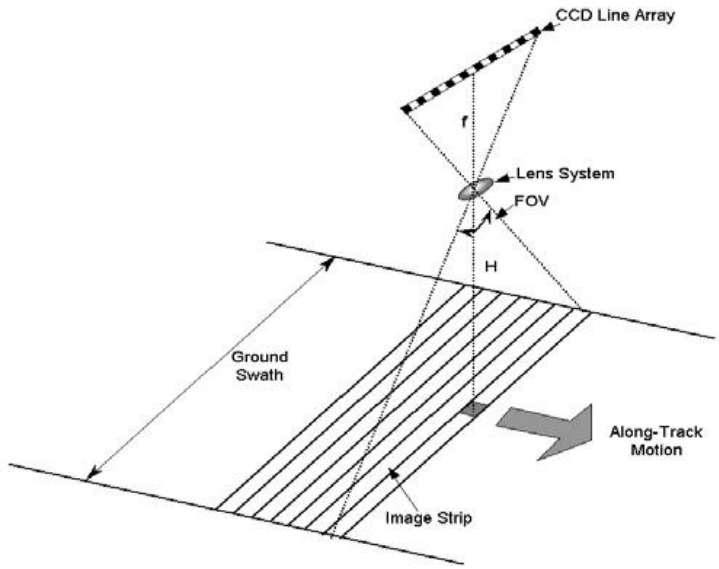


Figure 11: Working principle of Pushbroom Cameras. "*Cross-track coverage is determined by the number of detectors in the focal plane and the field-of view (FOV) of the lens system. The forward motion of the platform produces along track coverage.*" [Zhang et al., 2005]

The advantages of the new digital aerial cameras over analog cameras are fairly obvious and quite well known – chiefly that there is no film! This makes the handling of these images very convenient.

Data quantities can increase considerably without adding costs. That means that the number of images per project can be increased and thus vastly improve the redundancy of taken measurements. The traditional forward-/sidelap percentages at 60/20 can be increased to 90/60. This higher overlaps result in higher data quantities and require advanced computing techniques.

Moore's law is a welcome factor that promises that over the next 10 years, computing will improve its price-performance ratio by a factor of a 100. Image storage, retrieval, transmission, processing and information extraction are currently feasible with such increased data rates, and this feasibility will improve dramatically over the coming years [Leberl et al., 2004].

While traditional film photogrammetry shows one object point on only 2 images, with a digital system one oblject point can be visible in 10 images, with no extra cost for the images. Another dramatic improvement was the sensor technology. Digital aerial images often have approximately 13 bits of radiometric range, analog aerial images 7 to 8 bits per pixel. This implicates that also areas in an image with low contrast can be analyzed thanks to the

enhanced radiometric range of the RGB and infrared values. This is crucial for example for areas with intense shadows that occur quite often in dense urban scenes.

Table 6 summarizes the advantages of digital aerial cameras over analog cameras. *"The transition from minimizing the number of film photos to maximizing the robustness of automation by a free definition of the number of images can be a major shift in the photogrammetric value system, and can be called a „paradigm shift""* [Leberl et al., 2004]. The combination of these improvements in image quality and overlap has favorable consequences for the degree of automation for further processing steps; *"therefore, there is no more reason to use analog photos instead of original digital images"* [Jacobsen, 2010].

ECONOMY	*GEOMETRIC ACCURACY*
No consumables such as film processing	No image shrinking or other deformations
No scanning	Inner orientation is system-inherently available
No cost color and color IR	Identical duplicates
No cost of duplication	Absence of grain results in higher accuracy
Reduced manual labor	Radiometric range improves accuracy
IMAGE QUALITY	*WORK FLOW*
Note: 12 bits/pixel; No grain noise	In-flight quality control
Better matching accuracy (2.5 x film)	Automation with less manual intervention
More flying days in marginal weather	Improved manual interpretability (RGB & IR)
Better interpretability	No more film management
Better success in automated procedures	Intra-/ Internet- enabled archiving and cataloging
	Camera maintenance via Internet
REDUNDANCY IN MEASUREMENTS	
Note: Higher forward overlaps at no extra costs	
Better success in automated procedures	
Superior automated DEM generation	
Fewer occluded areas in urban projects	

Table 6: List of advantages of digital sensing over film cameras in photogrammetry [Leberl et al., 2004]

We will agree that considerable progress to characterize properties is available from information contained in aerial photographs. The accuracy and resolution of digital image acquisition systems has strongly improved during the last years. Table 7 shows all large format digital cameras that are available on the market and their key parameters. For our work, the most interesting parameters are the angle of view cross and along track that are crucial for the semantic interpretation of building facades.

Camera	pixels	Pixel size [µm]	f [mm]	frame rate [sec]	Pan/ MS	Angle of view cross / along track
DMC	7680 13824	12	120	2,1	4,6:1	69.3°/ 42°
UltraCam D	7500 11500	9,0	100	1	2,9:1	55°/37°
UltraCam X	9420 11310	7,2	100	1,6	2,9:1	55°/37°
UltraCamXp	14430 17310	6,0	100	2	3,0:1	55°/37°
UltraCamXp-WA	14430 17310	6,0	70	2	3,0:1	73°/52°
UltraCam Eagle	20010 13080	5,2	80 210	1,8	3,0:1	66°/46°
ADS80	12000	6,5	62	1000 lines/sec	1:1	64°/0°
JAS 150S	12000	6,5	150	800 lines/sec	1:1	75,5°/29,1
DMC II 140	11200 12096	7,2	92	2	2,0:1	50,7°/43,3°
DMC II 230	14400 15104	5,6	92	1,7	2,5:1	50,7°/46,6°
DMC II 250	14656 17216	5,6	110	1,7	3,2:1	45,5°/38,6°

Table 7: Technical parameters of most common Large Format digital aerial cameras [Jacobsen, 2011]

Redundancy

To be able to use these images for the semantic interpretation of real properties and buildings, we need massively redundant image information. With an 80% overlap in flight direction and 60% across flight direction every point is represented in 10 images, with a 90/60 overlap even in 20 images. In contrast, traditional analog photogrammetry maps each object point in only two images. The possibility to map object points in ten or more observations improves the measurement accuracy by a factor of six [Ladstätter et al., 2008]. This redundant image information can be used to minimize occlusions, to improve geometric accuracy, to capture all facades, to eliminate gross errors, to densify the Digital Surface Model etc. [Leberl et al., 2009a] so that one can determine the geometrical figure of every building with an accuracy of a few centimeters. This increase in redundant image data also leads to geometric and radiometric resolution, which is crucial for semantic image characterization because it enables that any characterization algorithm can rely on multiple analysis results that can either reinforce or cancel one another.

Figure 12 (left) Ground coverage of the UltraCam series; (right) Photograph showing UltraCamX by Vexcel Imaging

4.3. Satellite Imagery

Satellite Imagery and vertical aerial imagery are two related technologies that produce nadir images. The major difference is that traditional aerial images are taken from planes flying in the troposphere and stratosphere and satellite images are taken from satellites that orbit the Earth in the exosphere.

For the last decades, satellite imagery was considered to be the main data source for land cover and land use classification for larger areas as well as for the generation of coarse orthophotos, and there is still a lot of research going on in this field. The continuously reducing pixel sizes of satellite images are the driving force in this field. These developments make satellite imagery more and more interesting for image analysis like building detection and characterization. Since the launch of Geo-Eye-1, with a Ground Sampling Distance of 50cm, it is feasible to create coarse 3D models of man-made structures and there is a lot of effort invested in exploring new opportunities that this development offers.

However, even if the resolution of digital satellite images is improving over the last couple of years the main drawback of this technology is the insufficient geometric redundancy (mainly suppressed by a projection near to the orthographic ones) [Kluckner et al., 2009]. For interested readers in the next section we discuss the history of satellite imagery.

History of Satellite Imagery

World War II brought tremendous growth and recognition to the field of aerial photography. *"In 1938, the Chief of the German General Staff (General Werner von Fritsch) made a prophetic*

statement when he said: The nation with the best photo reconnaissance will win the next war."[7] The start of satellite imagery was in 1946, 8 years after General Werner´s statement. During the Cold War, the United States used U-2 planes to take photos of suspected Intercontinental Ballistic Missiles (ICBM) testing sites and air bases in the Soviet Union flying over 70000 feet. When on May 1, 1960 a U-2 plane was shot down by the Soviets during a reconnaissance mission, people realized that it was essential to take images from higher altitudes to look at conditions on the Earth's surface – the birth of "Remote Sensing"[9].

During the 1960's, the United States started space reconnaissance by taking images from Earth orbiting satellites, CORONA. The Corona program was a series of American strategic reconnaissance satellites produced and operated by the Central Intelligence Agency Directorate of Science & Technology with substantial assistance from the U.S. Air Force. The Corona satellites were used for photographic surveillance of the Soviet Union and the People's Republic of China. Until 1972, over 800000 images were acquired in 121 launches of these satellites with a ground resolution up to 1.5 meters [Corona, 1988].

In 1972, the National Aeronautics and Space Administration (NASA) started the ERTS-1 satellite (later renamed Landsat 1), which was the first satellite designed to study and monitor the Earth's Surface. "*Although space imagery from optical sensors has been available to civilian users on a regular basis ever since the successful launch of the first Landsat in 1972, it has attracted little interest among those photogrammetrists concerned with topographic mapping. This has been due both to the monoscopic nature of most space imagery and to its relatively low ground resolution. This meant that it was seriously deficient in terms of the features and content that are required even for basic small-scale mapping. Furthermore, until the advent of SPOT in 1986, the stereo-coverage required for the generation of accurate three-dimensional coordinate data was limited to that acquired by photographic cameras during manned space flights of short duration. These included the Spacelab missions between 1973 and 1975 and the individual Space Shuttle missions carrying the ESA Metric Camera (MC) and the NASA Large Format Camera (LFC) dating from the early 1980s*" [Petrie, 1998].

Until the mid 1990's, Remote Sensing was largely a government-led activity. "*The first opportunity occurred with the emergence of commercial remote sensing satellites (such as DigitalGlobe and GeoEye). The second opportunity was driven by consumer adoption of online mapping (such as MapQuest, Google Maps, Yahoo Maps, and Bing Maps), its appetite for high-resolution imagery through consumer-centric websites such as Zillow,*"[10] *and the emergence of personal navigation devices and location-based mobile apps.* The commercial satellite Geo-Eye-1 provides imagery with a pixel size of 50cm (see Figure 13) and its successor Geo-Eye-2 that is planned to launch in early 2013, with a resolution of 25cm, will open new applications using satellite imagery. The detailed characteristics of remote sensing satellites and the imagery acquired by them are set out in Table 8.

[9] http://www.oneonta.edu/faculty/baumanpr/geosat2/RS%20History%20I/RS-History-Part-1.htm
[10] http://www.grss-ieee.org/remote-sensing-history-1990-2010-%E2%80%93-the-evolving-private-sector-role/

Satellite	*Orbital Height (km)*	*Ground Coverage (km)*	*Ground Pixel (m)*
SPOT	822	60*60	10
IRS-1C/D	817	70*70	6
EarlyBird	475	6*6	3
QuickBird	600	22*22	1
IKONOS-1	680	11*11	1
OrbView3	460	8*8	1
EROS-A	480	12.2*12.5	2
EROS-B	600	16*16	1

Table 8: Characteristics of main sensors for Satellite Imagery [Petrie, 1999]

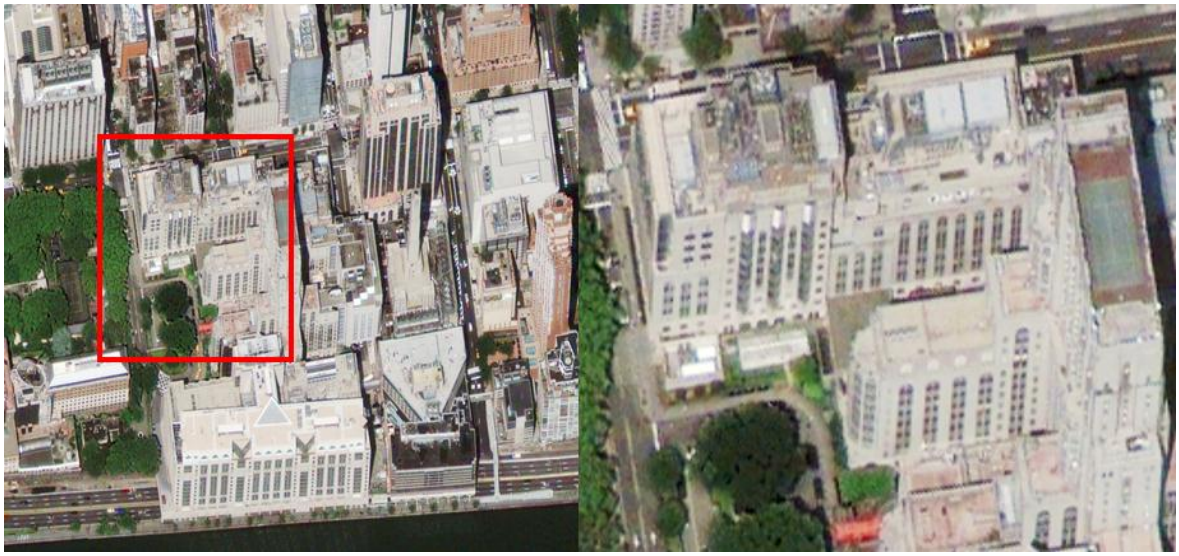


Figure 13: Satellite Image of Roosevelt Island, New York (GeoEye-1 satellite GSD 50cm; 2009) [from www.Geoeye.com[11]]

4.4. Oblique Aerial Images

There exists a need for simple measurements in urban spaces by non-expert users in municipal administrations. This combines with the appeal of a bird's eye view of a neighborhood and has led to the introduction of oblique aerial photography [Petrie, 2009]. In contrast to traditional vertical aerial images, the angle of view of oblique photographs lies between 35° and 65° [Prandi., 2008]. These oblique images are difficult to obtain using standard aerial cameras.

[11] http://www.geoeye.com/CorpSite/gallery/Default.aspx?gid=7

Therefore, to fully exploit the information from the oblique perspective, a minimum of four images from all sides are necessary. Such aerial images are more descriptive than vertical photographs, but are geometrically/photogrammetrically more difficult to process. By using the sloped image plane buildings are not photographed vertically from above, but from the side. This fact makes oblique aerial images the standard source of images for the characterization of urban structures from the sky. Thus, we pose the following question: Can one analyze facade details of buildings from overlapping vertical and oblique aerial photography, and at what level of completeness and accuracy? We will clarify this question in chapter 7. We will discuss there which of these two technologies is the most efficient and robust method for this purpose. Before we do that we want to take a closer look at the history and the technologies behind these special kinds of images.

History of Oblique Aerial Images

Oblique aerial images have a long tradition in the military sector where they have long been standard for reconnaissance purposes by taking advantage of their larger viewing angles. First, oblique images from planes were captured in the 1920's. One of the first cameras that took oblique images was the aerial camera K-3, designed by Sherman Fairchild that was mounted on an airplane to take high altitude oblique shots. As technology advanced, more and more lenses were added to the camera system. James Bagley of the US Army Corps of Engineers developed in 1917 a 3-lens oblique camera that was able to record three exposures simultaneously on one roll of film. Fairchild developed the T-3 5-lens camera around 1926 where 5 separate cameras were linked together. It consisted of one vertical and four oblique cameras that recorded images simultaneously[12] and produced "Maltese Cross" coverage of the ground. These cameras were extensively used during the Second World War.

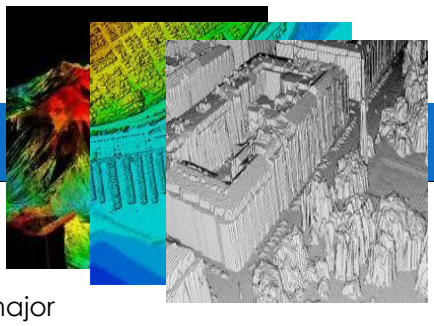
Revival of Oblique Imagery

Recent developments in digital photography and accurate navigation equipment have brought a new lease on life to oblique imagery. For many surveying and mapping companies, oblique imagery has traditionally been a side line business to the acquisition of vertical photography, providing side views of a location with little potential for measurement or real integration with GIS data. Traditional oblique photography has been either artistic to show a landscape or location-specific to support a particular application [Simmons et al., 2007]. This has dramatically changed within the last couple of years and currently there is a huge demand for oblique imagery. In this current revival of the old idea, now implemented using digital instead of film cameras, the target markets are law enforcement agencies and emergency (fire and ambulance) services. Apparently these organizations have found that oblique images are more easily analyzed by non-expert users, who also consider the information about building use and structure offered by oblique images to be useful for their operations. A recent application for this type of imagery has been in support of the location aware Internet, so far mainly in www.bing.com/maps, also in the regional site www.klicktel.de.

Currently, the market demands geometrically accurate images for texture mapping of 3D models [Frueh et al., 2004b; Grenzdörffer et al., 2007; Wang et al., 2011;Karbo et al.,

[12][http://www.nasm.si.edu/exhibitions/lae/html/sky_recon2.htm]

2009]and to measure distances and heights [Höhle, 2007]. Therefore, digital medium format cameras are coupled with GPS/INS for more precise georeferencing. Recently, there is developing research in the field of reconstructing and characterizing buildings from oblique images [Müller et al., 2007;Panday, 2011; Xiao et al., 2010].



4.5. Aerial Laser Scanning (LiDAR)

LiDAR, an acronym for Light Detection and Ranging, is an optical remote sensing technology that is used to measure distances. LiDAR point clouds have conquered a major position ever since point clouds have become a mapping data product in its own right [Toth, 2009]. In the last decade, LiDAR has become a valuable data source for urban data acquisition, and most of the state of the art methods for 3D building reconstruction and characterization are based on airborne LiDAR (see chapter 2).

History of LiDAR

The word LiDAR first appeared in the Middleton and Spilhausa "Wheater Instruments" in 1953.[13] At this time, it was used as light sources, conventional or pulsed lamps with high-speed shutter, which formed short pulse for mapping particles in the atmosphere[14]. In 1969, a laser rangefinder and a target mounted on Apollo-11 were used to measure the distance from Earth to Moon. But only by the development of the Global Positioning Systems (GPS) in the 1980s, allowing the precise positioning of aircraft, opened up applications to moving sensors. The ultimate breakthrough of LiDAR came in the 1990's during the telecommunications boom [BC-CARMS, 2006]. With the improvements of the IMU and the resultant increase in accuracies, the LiDAR technology was able to satisfy the big demand for mass production of high accuracy digital elevation models (DEMs), digital terrain models (DTMs) and triangulated irregular networks (TINs).

Working Principle

But how does LiDAR work? The principle behind LiDAR is really quite simple (see Figure 14). The LiDAR sensor essentially records the difference in time between the signal being emitted and received from a given point, very much like a conventional survey instrument.

[BC-CARMS, 2006] describe the LiDAR working principle in the following way: "*There is typically a LiDAR laser sensor, which is precision mounted to the underside of an aircraft, and which transmits or pulses a narrow laser beam towards the earth as the aircraft flies. A receiver additionally affixed to the aircraft receives the reflection of these pulses as they bounce off the earth below back to the aircraft. Most LiDAR systems use a scanning mirror to generate a swath of light pulses. Swath width depends on the mirror's angle of oscillation, and ground-point density depends on factors such as aircraft speed and mirror oscillation rate. Ranges are determined by computing the amount of time it takes light to leave an airplane, travel on the ground and*

[13][http://xipar.info/historylidar.html]
[14] [www.csc.noaa.gov]

return to the sensor. A sensing unit's precise position and attitude, instantaneous mirror angle and the collected ranges are used to calculate 3-D positions of terrain points. As many as 100,000 positions or "mass points" can be captured every second. The LiDAR sensor essentially records the difference in time between the signal being emitted and received from a given point, very much like a conventional survey instrument. The LiDAR data is coupled with additional precise positioning information gathered by on board Global Positioning Instruments (GPS) and other Inertial Navigation Systems (INS). Once the total information volume is stored and processed, the resulting product is an extremely accurate "x.y.z." for every position scanned on the ground."

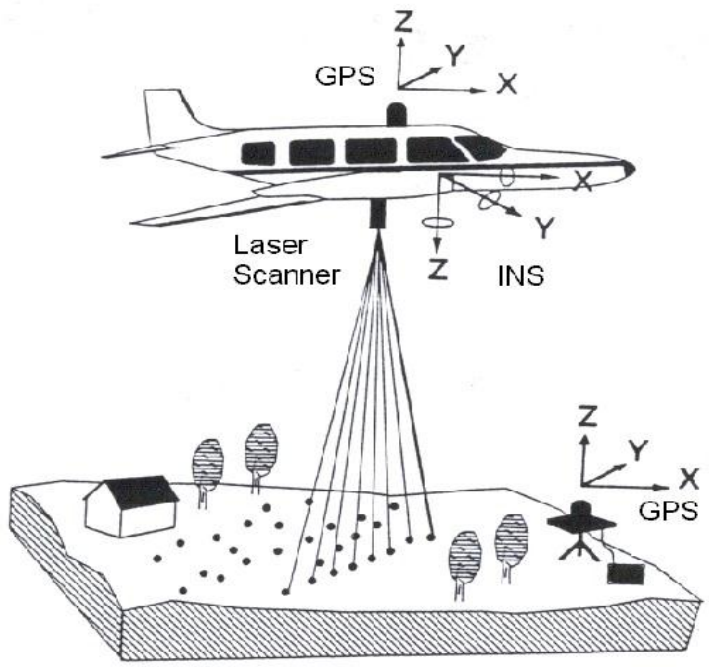
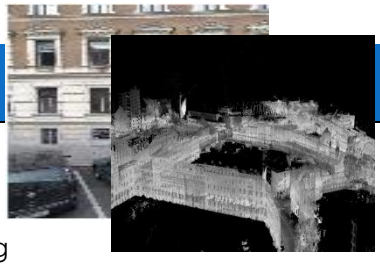


Figure 14: LiDAR Working Principle [BC-CARMS, 2006]

Modern LiDAR systems are capable of emitting 400000 pulses per second[15]. Airborne LiDAR systems conventionally measure one or more ranges to objects present in the trajectory of the laser pulse, as well as strength of backscattered signal. Typically one (first or last), two (first and last) or a few (up to four) return signals are recorded for each pulse. The strength of the return signal gives a clue concerning the type of object. Most LiDAR systems developed for land applications use near infrared laser, so high-energy returns indicate that the pulse has been reflected from trees, meadows or other vegetation. Recording of the first and last pulse makes it possible to distinguish between canopy and ground surface. The signature of the return signal is determined by interaction between laser beam and objects present in the footprint, and can be inferred from capturing the full waveform. In the full waveform approach, the entire backscattered signal is sampled at equal intervals over a hundred nanoseconds depending on object height.

[15][http://www.gim-international.com/issues/articles/id1667-Airborne_Lidar_Scanners.html]

4.6. Street Level Images and Lasers



Mobile Mapping Systems have slowly developed over the last 20 years. In the beginning, companies developed in-house systems to collect data on road infrastructure or building facades. The desire for a human scale pedestrian experience of urban spaces on the Internet, in combination with an immersion in the 3D urban World, have led to the advent of street side imagery as a major source for urban mapping data. The initial impetus was for industrial systems on cars to drive through all city streets to collect images of facades. A small industry has emerged with cars and vans carrying multiple cameras on their roof. The breakthrough came in 2007 when Google launched Google Street View, Microsoft launched StreetSide View and Navteq and TeleAtlas were founded (see Figure 15). The existing technology was adopted on a large scale and a substantial fleet of mobile mapping vehicles emerged. Currently, these mobile mapping vehicles are being used worldwide for their imaging and mapping operations. This led to a rapid improvement of the technology that led to a boom in this field.

The high resolution (up to 1cm) of these systems makes it very attractive to reconstruct and characterize building facades and there is a lot a research in this field. Imaging aims at the collection of panoramic views employs multiple centrally perspective cameras, such as the LadyBug (www.ptgrey.com). Cyclomedia (www.cyclomedia.com) pioneered the commercial use of 360° panoramic views of urban spaces and has been relying on fish-eye imaging since this approach reduces the number of individual images being collected and therefore simplifies geometric processing. Panoramic views combining a conventional camera with a spherical mirror carry this simplification even further, in which a single image can present the entire hemisphere, is a frequently proposed approach for indoor imaging that uses catadioptric cameras.

The cost of imaging is a function of the kilometers driven by the vehicle. The mid-sized city with 250,000 inhabitants may have 1000 km of streets that will nominally be driven at a velocity of 30 km/h and thus take a net 33 hours. However, traffic, one-way streets and weather constraints (no rain) will extend this to a velocity of well under 30 km/h. The total cost of coverage may not be that different from an aerial mission [Leberl et al., 2012].

Figure 15: Example of Mobile Mapping System mounted on a car [taken fromwww.navteq.com]

In the last couple of years, street side images have become the workhorse for urban modeling in Google Maps and Bing Maps. These images are presented for visual orientation, but are not interpreted to describe the street scenes shown in the images so far.

Working Principle

A Mobile Mapping Device consists in general of three main parts: The digital camera devices, the laser ranging and scanning devices as well as the positioning components.

Concerning the camera device, digital frame cameras are used with a resolution of 1 to 3 Megapixels and a frame rate between 7 to 15 frames per second. The cameras are placed to produce 360 degree panoramic images in the horizontal plane with one camera pointing vertically upwards. These specifications ensure a continuous coverage of a street.

The second component is the laser ranging and scanning devices. The 3D laser scanners that are used in surveying are not suitable for moving vehicles because their scanning operations need too long. Instead, for this purpose 2D laser scanner are used that can rapidly acquire range or elevation profiles comprising the distance and angular values measured within a single 2D plane. There are several providers on the market like SICK, LASE GmbH, Ibeo, Riegl Topcon and Optech that sell such products.

The third component is the positioning devices such as the GPS and the IMU that are necessary to determine the absolute location and orientation of the moving vehicle and its mapping system. Additionally, most systems also use a distance measuring instrument (Odometer) to improve the overall accuracy.

4.7. Crowd Sourced Image Data & Community Photo Collections

A sudden development is the use of Community Photo Collections, as documented in the widely published "Rome-in-a-Day" paradigm [Snavely et al., 2008; Snavely et al., 2010; Agarwal et al., 2009]. The idea is to have the public collect images of street canyons and buildings, upload those with some geo-positions from GPS into a system such as FLICKR, and then process these photo collections into 3D urban models. Figure 16 illustrates this approach. The source data are available at no cost.

A mid-sized city with 250.000 inhabitants may consist of perhaps 75.000 structures, thus buildings, sheds, garages, industrial buildings etc. If every building facade is photographed by users onto perhaps 20 photographs, the total amount of data for such a city would encompass 5 million images. When each image has a resolution of 10Mpixels leading to 30 MB, the data set would have approximately 150 Tbytes [Leberl et al., 2012]. This is a huge amount of data, and to handle it one would need special data structures and high speed computers. However, this data combined with street side imagery could be used for urban scene characterization and in our case for building characterization.

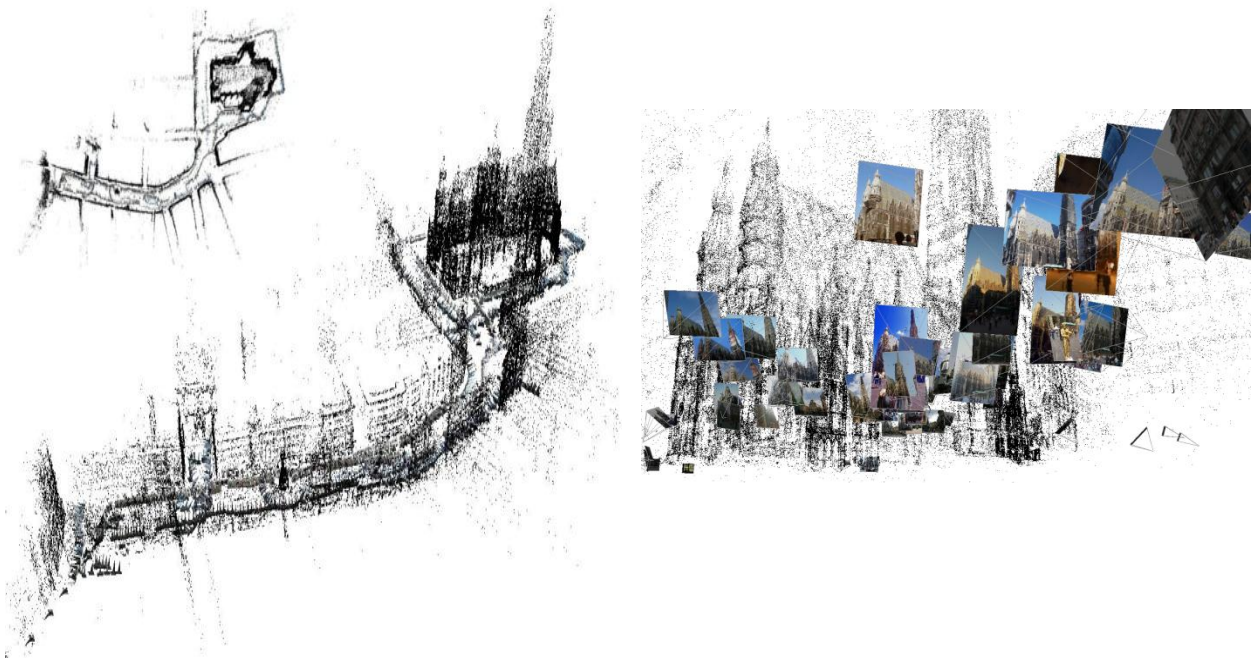


Figure 16: Example generated sparse point cloud using community photo collection showing St Stephans cathedral and Kärntnerstrasse in Vienna [Irschara et al., 2007]

4.8. Unmanned Aerial Vehicles (MAVs)

Since many years there exist small remotely controlled helicopters and small remotely controlled fixed wing planes. However, until now most of these systems suffered from a lack of accuracy and low image resolution. This changed in recent years due to the increase of the camera resolution and new algorithms in the field of image processing like dense matching [Wendel et al., 2011].

While their dimensions of up to 1m by 1m are reasonable small to allow a simple handling, they are still able to carry payloads of 0.5 to 1kg. Typical systems consist of a camera, some computing capability and data storage, as well as GPS and INS.

One may consider a system with remotely controlled Micro Aerial Vehicles (MAVs) flying along the streets of an urban area collecting imagery. A single MAV can today collect images during a single aerial mission for perhaps 15 minutes recording 5GB of image data that can be processed to generate 3D point clouds as illustrated in Figure 17. By the use of multiple MAVs, the effort of image collection for an urban area might be feasible. We estimated that street side data over a distance of 15km could be acquired within an hour using one helicopter [Leberl et al., 2012]. So far, the major drawbacks of this technology are undefined legal restrictions in urban areas. Because of this, it is not easy to get approval by authorities for flying over urban canyons and its people.

Figure 17: (left) example of an UAV; (middle) collected images of an atrium scene (right) computed sparse point cloud from these images [Leberl et al., 2012]

4.9. Catastre

Since a "property boundary" is a legal concept, it is not typically visible in the field and from the air. Also image segmentation algorithms cannot properly distinguish between buildings when they are physically attached to one another. It will be the rare exception that attached buildings can be separated from aerial imagery, for example if the roof styles differ, building heights vary or the colors of the roofing tiles differ. Moreover, boundaries from aerial images are often not evident. Thus, the introduction of the cadastral map is the only obvious answer.

*A **cadastre** is a comprehensive register of the meet's and bounds real property of a country, and commonly includes details of the ownership, the tenure, the precise location (some include GPS coordinates), the dimensions (and area), the cultivations if rural and the value of individual parcels of land.*[16]

In the cadastre all plots and physical structures of a property are described and mapped according to their position, use and size. Cadastral information has a long history that is reflected in a need to consider its national circumstances and historical development to avoid negative surprises. The currently valid Vermessungsverordnung in Austria became effective on the 1st of December, 1994 and standardizes the accuracy values of control and border points that have to be observed. This regulation considers historical circumstances concerning the determination of reference points (measurements without electronic distance measurements or GPS) and defines based on the average accuracy of the official reference points (± 5 to 7 cm) a tolerance for border points (± 15 cm) which may not be exceeded.[17] The mentioned accuracies have to be considered when combining the cadastral information with the aerial images of a certain region. The next section should give a short overview about the history of the Austrian cadastre.

The first attempt to introduce a registration system was in 1718 for the Duchy of Milan. The process entailed mapping and surveying all buildings and parcels. In 1817, the Franziscan Cadastre was introduced to systematically register all parcels of the whole country based on

[16] Wikipedia (cadastre; last accessed 12/2011)
[17] http://www.meingrundstueck.at/Lexikon/L-Genau/l-genau.html

field survey. The main goal was to introduce a taxation based on achievable profit from real estate based on parcels that treats all land owners equally. At that time, only the land and not the buildings of a property were taxed; therefore, buildings were measured with a lower accuracy [Lego, 1967]. This fact means that in Austria, for example, the use of property boundaries for the interpretation of an aerial image makes sense but the use of building information out of the cadastre has to be handled with care.

The cadastral information addresses property boundaries, and not necessarily building footprints, although such building data may be provided from the cadastral sources, yet with some uncertainty and oftentimes dated. The cadastral accuracy is being quoted at ± 10cm which is at the range of the aerial photography's pixel size and thus sufficient for the purpose of characterizing real properties [Navratil et al., 2010].

However, the accuracy and the historical development of the Austrian land register can be considered to be at a rather highly developed level not found in many countries. There is little we can do when dealing with countries where the accuracy of the cadastral data doesn't fulfill the necessary requirements or in the worst case scenario the data is not available. In the next section we present data that can be used in such a case.

4.10. Optional Street Network and Address Data

A solution to solve the problem, if cadastral data is not available or has been corrupted, could be the use of digital street maps. This data could be used to not only replace cadastral data, but also to add additional information to the database like traffic related information. Car navigation has been the driver for the global development of these maps. They also support location awareness of Internet searches. One of the biggest providers of this data is OpenStreetMap.org. The OpenStreetMap project was started in 2004 by Steve Coast. The goal of OpenStreetMap is to generate and distribute free geospatial data for the masses. In the beginning, the initial map data were collected from scratch by volunteers performing systematic ground surveys using a handheld GPS unit, a notebook, a digital camera or a voice recorder.[18]More recently, aerial imagery and other data sources from commercial and government sources have greatly increased the speed of this work. Today, the OpenStreetMap database contains more than 2.5 billion GPS points, 107 million roads and 1.1 million relations (State: August 2011).[19]

Since the vector format of the street maps is not available for free, it is at times re-engineered from the raster display off the Internet [Chiang et al., 2008].

Address information is encoded in the form of a point and its geographic latitude and longitude. This is valuable when attaching alphanumeric information to a property. Again, this source is of elevated importance when one needs to cope with the absence of a cadastre. Between the street map, address data and the orthophoto one will have to devise a strategy for approximating cadastral information. Figure 18 illustrates the scarcity of a street map for one test area.

[18] [http://en.wikipedia.org/wiki/OpenStreetMap]
[19][http://www.openstreetmap.org/stats/data_stats.html]

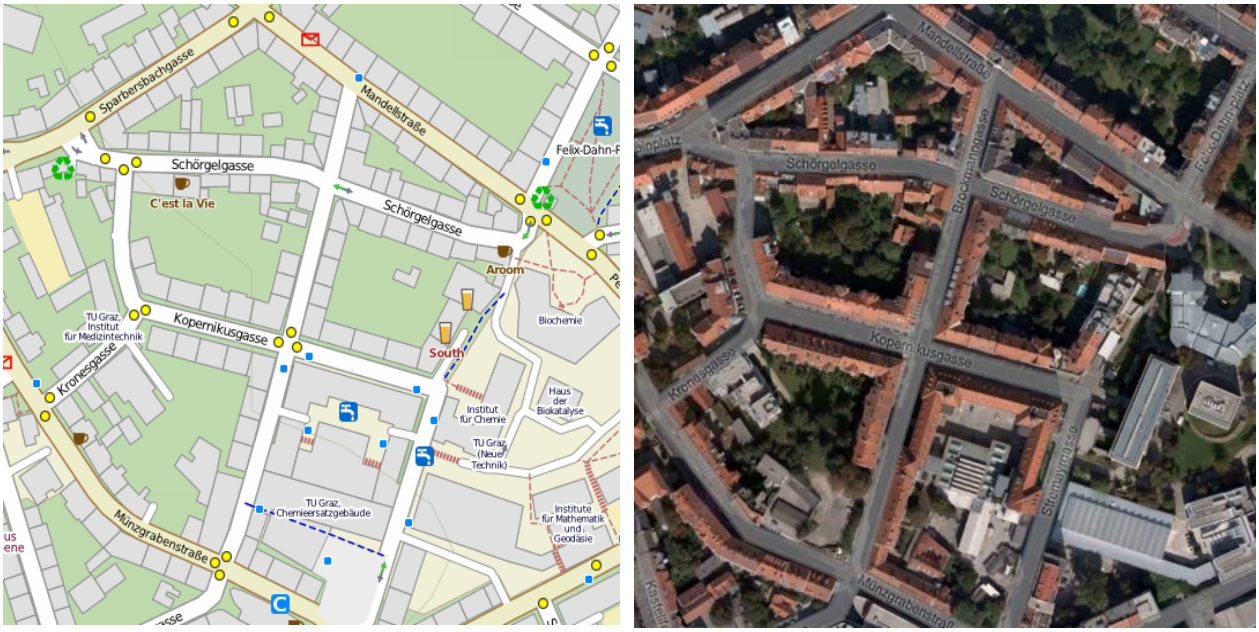


Figure 18: Street layer for the area in Figure 6 obtained from the website http://www.openstreetmap.org/, Nov-2011). The raster GSD has been resampled to fit the aerial imagery sources at a GSD of 10 cm. To the right is a "hybrid" view of the street layer superimposed onto the orthophoto.

4.11. Derived Products

To be able to characterize real properties we need additional datasets like the Digital Terrain Models (DTM), Digital Surface Models (DSM), True orthophotos and classification images. In the next section, we will discuss the generation of these data products using vertical aerial images in more detail and the necessary processing steps.

Aerial Triangulation

The fundamental principle used by photogrammetry is triangulation. By taking photographs from at least two different locations, so-called "lines of sight" can be developed from each camera to points on the object. By mathematically intersecting converging lines in space, the 3-dimensional coordinates of the points of interests (POI) can be determined and all images are transformed in a common coordinate system. When reconstructing 3D models from overlapping images, or if objects have to be recognized from different images, a sub-pixel match of the overlapping images is important to improve accuracy.

At first a POI extraction in each image has to be performed to find correspondences between two images. By the use of affine invariant area based matching the number of candidates is reduced. Furthermore, in order to fulfill the non-ambiguous criteria, only matches with a high distinctive score are retained. For getting robust results a back-matching is applied. This is done for all consecutive image pairs.

Finally, a bundle block adjustment calculates the relative orientation and incorporates GPS data, IMU data or ground control information.

Standard photogrammetric processing is being applied to such a block of digital photography using the UltraMap-AT processing system. The word "standard" is applicable in terms of the principle of photo-triangulation itself, but the procedure is far from standard when considering the data quantities and the level of automation [Reitinger, 2008]. Full automation is achieved first because of the high image overlaps; a second factor is the use of a much larger number of tie-points than traditional approaches have been using. The aerial triangulation process therefore results not only in a pose for each aerial photograph, but also in a sparse cloud of terrain points, since about 1000 tie-points are found per aerial photograph [Reitinger et al., 2008; Gruber et al., 2008b] as shown in Figure 19. The accuracy of the triangulation is in the sub-pixel range of ± 4 cm in planimetry and ± 6 cm in height [Scholz et al., 2009] using photographs at a GSD of 10cm. The sparse cloud of terrain points may be at 20,000 per sqkm. The result of this computation is the basis for the DSM, DTM, Classification Image and True Orthophoto generation. Depending on the accuracy of the aerial triangulation the generation of these products will fail or be successful

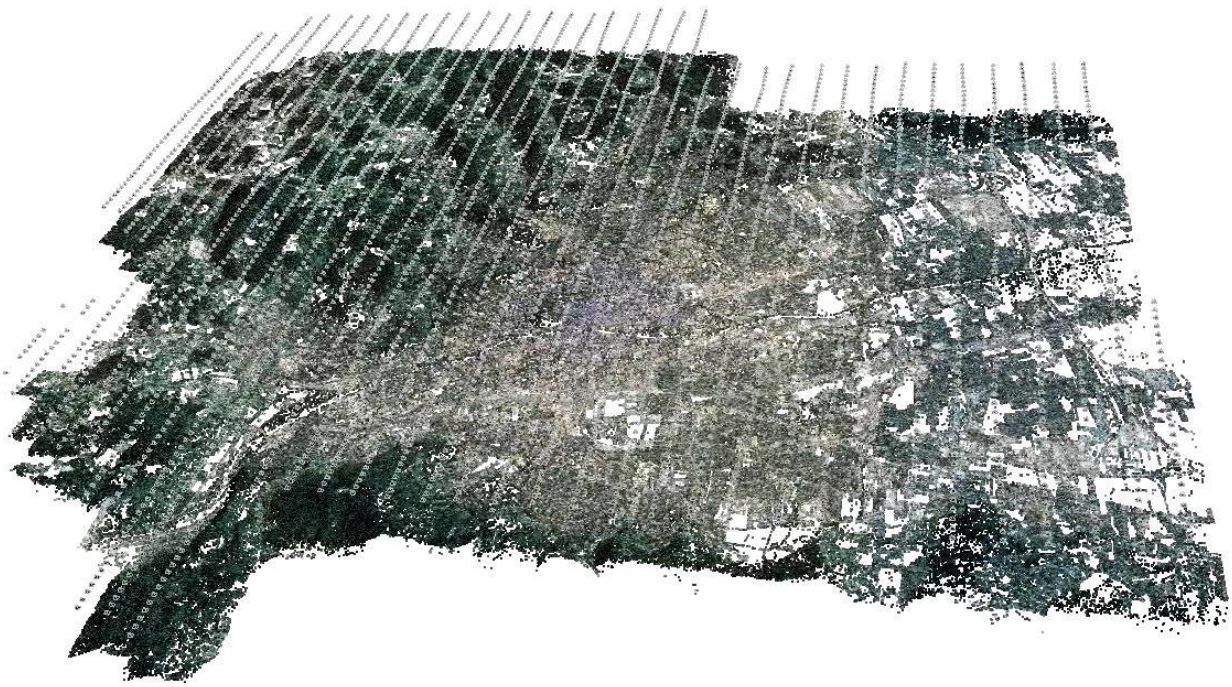


Figure 19: Aerial Traingulation (AT) result from 3000 aerial images covering an area of approximately 150 km² of Graz and surroundings [Irschara et al., 2007].

DSM and DTM Generation

The estimation of depth information from overlapping aerial images is essential and therefore a well-studied problem [Hartley et al., 2004; Kraus, 2004]. The input consists of highly overlapping triangulated aerial photographs. The depth information of a specified scene can be generated from at least two overlapping views, generally denoted as stereo image pair. Therefore,

corresponding points are detected in all images using e.g. correlation-based matching methods [Hartley et al., 2004]. The Digital Surface Model (DSM) is created by "dense matching" (see Figure 20). In the process, one develops point clouds from subsets of the overlapping images and then merges (fuses) the separately developed point clouds of a given area. The process is by [Klaus, 2007]. For a property classification the elevation data are being considered as input. Data used in our property-related experiments have been obtained from work at Microsoft Graz [Scholz et al., 2009]. The conversion of the surface model DSM into a Bald Earth Digital Terrain Model (DTM) is a post-process of dense matching and has been described by [Zebedin et al., 2006].

The postings of the DSM and DTM are here at 2 pixel intervals, and result in about 25 million points per sqkm, far denser than traditional photogrammetry rules would support. However, because the overlaps of 80% and 60% result in 10 images per ground point, one is observing a type of "super-resolution" that supports a very dense sampling of terrain points. It has been shown that the vertical accuracy of such DSM is in the range of about ± 0.5 pixels [Leberl et al., 2010a], provided the surface is well defined.

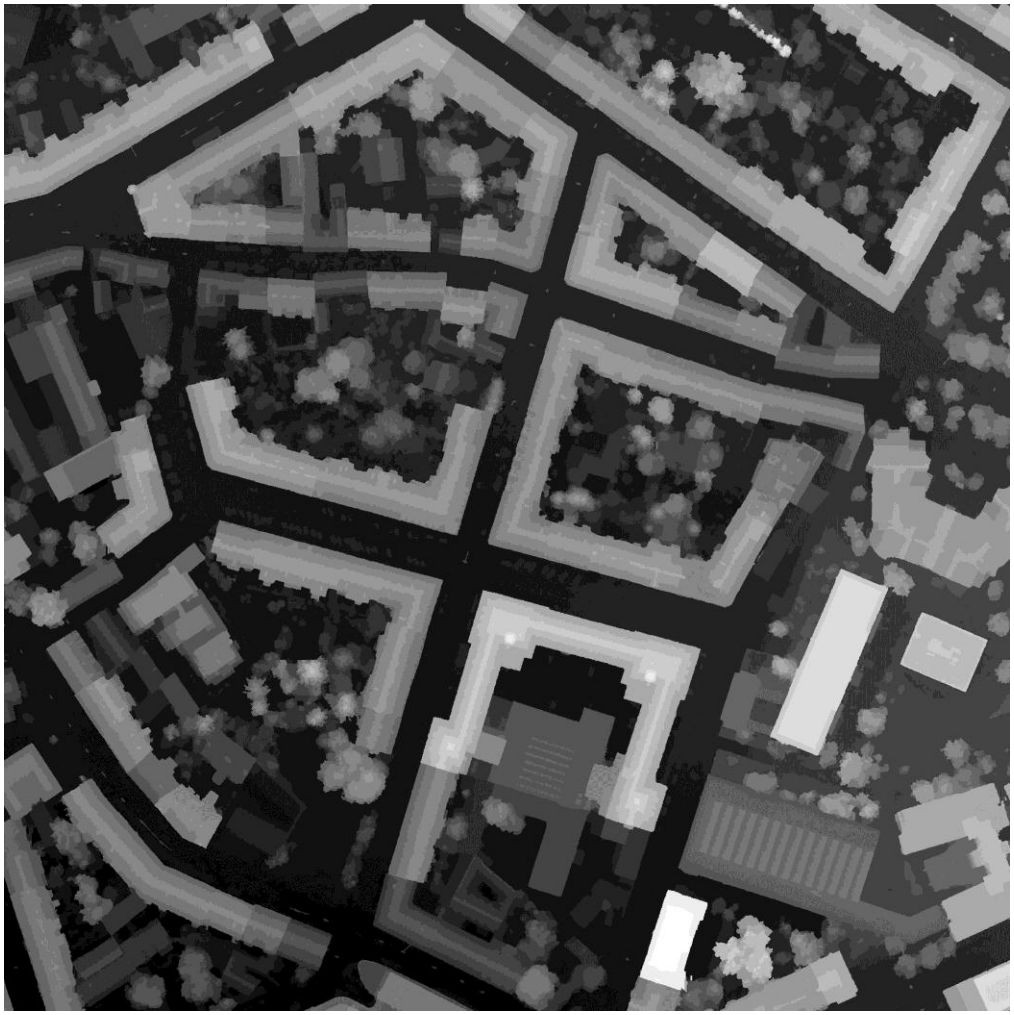


Figure 20: Digital Surface model for one test area from Graz data set (Vexcel Imaging)

Classification into Segmented Images

Any urban area of interest is being covered by multiple color aerial images. These can be subjected to an automated classification to develop information layers about the area. We consider these to be an input into our characterization procedures. Standard classifications of 4-channel digital aerial photography typically lead to 7 separate data layers: buildings; grass; trees; sealed surfaces; bare Earth; water; other objects ("unclassified"). The unclassified areas may show lamp posts, cars, buses, people etc. Classification and segmentation methods are topics of intense research. The classification approach used for this research has been introduced by [Gruber-Geymayer et al., 2006a; Zebedin et al., 2006]. [Kluckner et al., 2009] has proposed alternative novel methods with good results specifically interpreting urban scenes imaged by the UltraCam digital aerial camera.

[Gruber-Geymayer et al., 2006a] report typical classification accuracies in the range of 90%: this means that a 10% error is being made either by assigning a class in error or not assigning any class. Problems in this connection cause construction sites, undercrossings, embankments, cranes and bridges that are not buildings, to be classified as such. Further details for these steps are described in [Zebedin et al., 2006; Gruber-Geymayer et al., 2006b]. Figure 21 illustrates the result of this segmentation method (orange: buildings, dark green: trees, green and turquoise: vegetation, yellow: sealed surfaces, red: small high objects, blue: water).

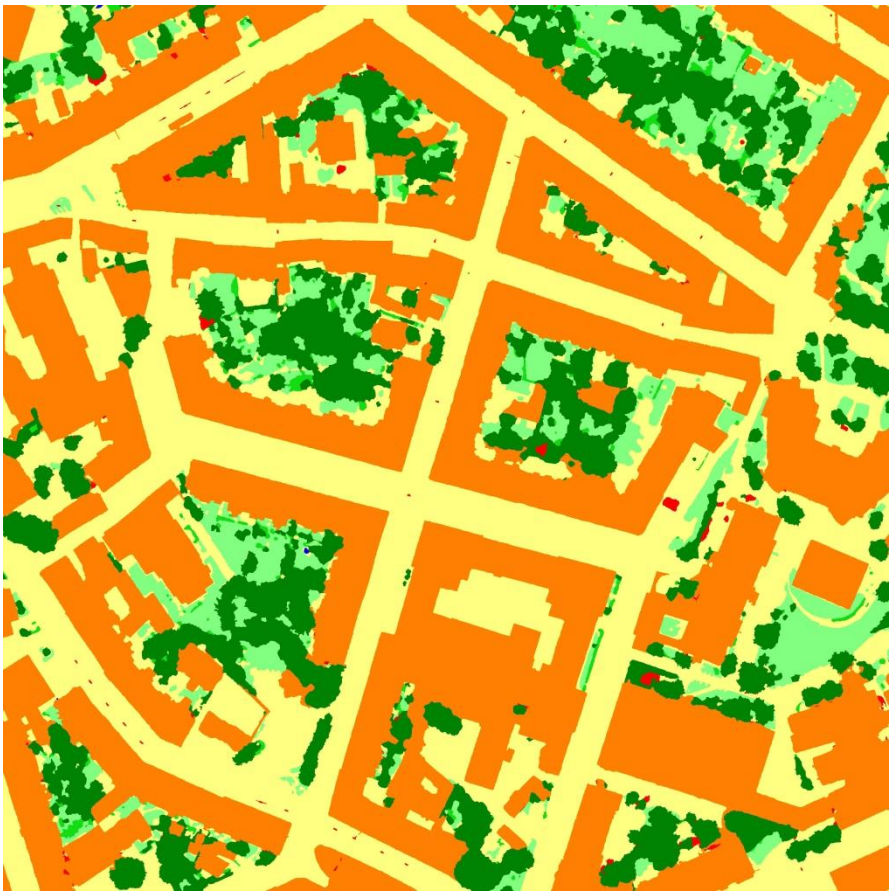


Figure 21: Classification Image for one test area from Graz data set (Vexcel Imaging)

In chapter 6, we have performed an evaluation of the classification method introduced by [Gruber-Geymayer et al., 2006b] with special interest on the building classification. We achieve a detection rate of 89% for the Graz test data set, which corresponds to the accuracies stated by the authors.

True Orthophoto Generation

True orthophotos have become a standard product of photogrammetric workflows. The DSM is the reference surface onto which each aerial photograph gets projected. The DSM and its associated photographic texture are then projected vertically into the XY-plane and result in what is denoted as a "true" orthophoto. In this data product, the buildings are only shown by their roofs and not by their facades.

In the current context, the true orthophoto is an auxiliary tool for visual inspection of intermediate results (see Figure 6). One could also use traditional orthophotos that have been projected onto the Bald Earth DTM instead of the DSM. For further details about true orthophoto generation see [Zebedin et al., 2006; Oda et al., 2004].

4.12. Experiments: LiDAR versus 3D Vision

The next chapter is reprinted from a joint work of [Leberl et al., 2010a]. The reprint of this paper as sub-chapter 4.12. occurs with the explicit approval of my co-authors in that paper. The defining capability of the stereo-photogrammetric method was to measure an object's geometry, in particular its 3D-shape [McGlone et al., 2004]. Since the mid-1990s, the geometric shape measurement by stereo has often been replaced by the direct measurement of point clouds by laser scanners, significantly reducing the role of photogrammetric technology. This erosion is documented by the growing density of technical and scientific publications and conferences on laser scanning versus photogrammetry [ILMF, 2009; DGPF, 2009; ASPRS, 2009]. Imagery serves in those cases merely to provide photographic color texture for orthophotos and for 3-dimensional object models created by lasers. Effectively, this reduces photogrammetry to a "dated" technology, made somewhat obsolete by new laser technology, and its application to the measurement of shape. It is thus of interest to obtain an understanding of the permanence of this development. To this end, we begin with a description of laser scanning capabilities as they relate to a modern photogrammetric system.

In the world of film imaging, a 3D point cloud came at the end of a long workflow that started in the air with heavy and costly rolls of environmentally fragile film, proceeded to the photo-lab for film development, went through scanning of the film images, then through a manually-supported aerial triangulation and achieved the point cloud in smaller patches by two-image stereo matches from pairs of images with 60% stereo overlap [McGlone et al., 2004]. These patches were being merged via the results of the triangulation so that a large area was being covered in a seamless assembly of patches representing a single point cloud. The time from taking the images to actually seeing the digital terrain model spanned weeks.

LiDAR View

In this world, the direct assembly of point clouds from laser scanners is appealing because the time from sensing to actually inspecting a surface model is reduced to almost zero: no film, no film-to-pixel-conversion, no matching. The point cloud is viewable as it gets downloaded from the sensor.

Laser scanning is predicated on the use of global positioning (GPS) and inertial guidance (IMU) to position and orient the sensor in a 3-D world coordinate system. This reliance on direct geo-positioning is seen as a second major advantage because there is no need for any type of triangulation and thus saving elapsed time and manual labor. However, also photogrammetric workflows often use GPS/IMU for the avoidance of an aerial triangulation and of the need for ground control points. Yet, total reliance on GPS/IMU does compromise the accuracy of the resulting stereo matches. Similarly, accuracy limitations of the GPS/IMU-positioning/orientation will cause geometric mismatches in the laser data along the overlapping seams of separately collected data strips. Those will need to be removed by some adjustment with manual interventions.

Aerial Imagery View

Dramatic potentials have emerged for advances in photogrammetric technology due to the transition from the hybrid approach characterized by the use of film, film scanners and digital replicas of the manual stereo process to the now fully digital workflow, characterized by the use of digital aerial cameras and fully automated workflows. Making these advantages work in everyday practice is a challenge currently faced by the photogrammetric industry. We have identified 6 separate areas of progress and innovation. Some of these still need to be developed for everyday practice. For example, multi-image matching for terrain modeling is not yet widely available, and early implementations most often do not warrant that designation since they oftentimes repeat a two-image stereo match and then fuse (average?) the multiple 2-image results.

Since its initial inception, photogrammetry was driven by minimizing the number of (film) images for any given project. Every additional image caused additional cost for materials and additional labor for dealing with the image, which resulted in yet another stereo model to be manually processed. A surface point was defined by 2 optical rays, representing 4 equations to solve for the 3 unknown coordinates x, y and z. These 4 equations for 3 unknowns have led to a photogrammetric workflow that hardly satisfied the surveyor's rule for reasonable redundancy. The transition to digital sensing did away with the extra cost for extra imagery. Automation does away with extra labor per image. Multi-view geometry does away with the idea that an additional image necessitates additional work with an additional stereo model [Hartley et al., 2004]. Images can now be produced at 80% forward overlap increasing the number of images per object point from 2.5 to 5, at absolutely no additional cost of acquisition. At 90% forward overlap, the images per object point within a flight line grows to 10. And by an increase of the side-lap from the traditional 20% to now 60%, the add-on cost will increase for the additional air time, but not for the increase in the number of images. The strategy increases the number of images per object point to 10 (at 80% overlap) or even 20 (at 90% overlap). The benefits are numerous – reduced occlusions, higher level of automation, reduced occurrence of

blunders/gross errors and therefore for manual editing, and finally an increase of geometric accuracy.

Comparison of Dense Matching Results from LiDAR and Aerial Imagery

Main advantage – Removal of Vegetation

A frequently made argument in favor of aerial LiDAR scanning addresses the ability of a LiDAR sensor to produce multiple reflections by one single pulse, creating a surface of the "bald Earth" and separately mapping the top of all objects on the bald Earth. This is the classical forestry application of LiDAR scanning [ILMF, 2009]. However, most photogrammetric software systems for the creation of digital surface models DSM offer a module for filtering of the bald Earth (DSM - DTM). We are not aware of any thorough studies that would show how well 3D-photogrammetry succeeds in separately mapping the bald Earth and its vertical objects. The wide-spread opinion is that 3D-photogrammetry is unable to deliver data that would be as useful as those from LiDAR.

Cost and Quality of Airborne LiDAR and Aerial Imagery

Aerial LiDAR point density may be in the range of 1 to 5 points per square meter. Aerial photogrammetry point density may at first be seen as defined by the 2D ground sampling distance (GSD), thus pixel size. At a 10 cm pixel, one would produce 100 points per square meter. But these are not 3D elevation values. The 3D sampling interval depends on the image overlap to achieve independent 3D matches. If 50% of the pixels were to get matched for independent 3D elevation values, one would achieve 50 points per square meter from 10 cm pixels. The difference between LIDAR and 3D vision is more distinct when one considers swath width and flight lines.

LIDAR	*DIGITAL PHOTOGRAMMETRY*
170 scans per second (190 kHz), 30° FOV;	GSD 25 cm
8 points per square meter	16 points per square meter
Flying height 750 meters	Flying height 4188 m
Aircraft speed 60 m/sec	*Aircraft speed 141 m/sec*
Strip width 403 m	Strip width 4328 m
20% side-lap between flight lines	60% side-lap between flight lines
Effective strip width at 322 m	*Effective strip width 1731 m*

Table 9: Project parameters for a laser system and a digital photogrammetric camera. Note the differences in strip width and aircraft speed, which cause the photogrammetric data collection to be almost 13 times more economical at these parameters. In addition, the higher number of photogrammetric points per m2 produces a better definition of the object shape along sharp edges. And the higher side-lap of the image-based method reduces the occurrence of occlusions. [Leberl et al., 2010a]

Table 9 presents some typical project values for a LiDAR Scanner flying at 750 meters above ground at 60 m/sec. The flight lines are 322 meters apart. An equivalent photogrammetric system is a digital camera flying at ~ 4,000 meters for a pixel size of 25 cm and at an aircraft speed of 140 m/sec. The flight lines are 1730 m apart. Under these assumptions in Table 9, we would have to be in the air 13 times longer with a laser sensor than with a camera to obtain comparable elevation results. First, the aircraft speed would be 2.5 times lower for the laser system than for the camera; second the laser swath width would be about 1/6th of the camera's. These differences in economy would be further aggravated by the laser's suffering from more occlusions than the camera because we did assume in Table 9 that the camera is flown with a 60% side-lap, and the laser with a 20% side-lap. The GSD at 25 cm, assumed in Table 9, is not the most popular with aerial imagery of urban spaces. More likely is the resolution at 15 cm and down to 8 cm per pixel. At those geometric resolutions the objects get defined at densities of 100 points per m^2.

Experiments

Aerial laser scanning and aerial 3D photogrammetry can be related via some experimental data collected by the University of Stuttgart over the Vaihingen test site [Cramer et al., 2009; Haala et al., 2009]. Figure 5 superimposes over a color image taken with the UltraCamXp digital aerial camera the laser point cloud, taken with the Leica ALS50 scanner. The laser points have been sampled at an interval of 45 cm across and 70cm along the flight path. This results in a formal density of 4 laser points per m^2. The underlying aerial photograph is at 8 cm per pixel. Figure 23 compares a building detail from the laser point cloud to the 3D cloud of photogrammetrically matched points. We have three accuracy observations. First is the photogrammetric aerial triangulation with a σo at ~ ±1 µm or at ± 1.1cm on the ground, using CCD elements with 6.2 µm. This needs to be related to the GPS/IMU-accuracy of the laser scanning system. We have no verification of that accuracy in the given experiment. Generally, one can find literature assigning the GPS/IMU-vertical position accuracy with a value in the range of ± 20 cm, but with best cases in the range of ± 3cm.

Second is the accuracy along discontinuities. This is very much defined by the point density; the dense photogrammetric matches were achieved at a 2-pixel interval and are thus at 25 points per m^2. The laser scanner resulted in 6 points per m^2. It is visually clear from Figure 23 how the roof line's definition is superior in the photogrammetric data set. Third is the vertical data noise. Ideally, this should be in the range of the triangulation's uncertainty when surface points are being computed from all images and when surface texture is good. That would support the expectation of a vertical noise in the range of ± 3 cm. However, taking the flat soccer field in Figure 22 and looking for the variation of the terrain elevations presents another measure of "noise". The photogrammetric data show an uncertainty of better than one pixel, at ± 7 cm. This is twice what the triangulation result would predict, and we assume that this is caused by a failure to take advantage of all images in the matching process. Additional research is needed to develop a better understanding of the effect of various approaches to multi-image elevation measurements. The same approach applied to the laser data produces an uncertainty of ± 2 cm.

Figure 22: Color aerial photograph taken with the UltraCamXp digital aerial camera, and superimposed the measurements from the Leica ALS50 laser scanner. Direction of flight is to the right. Laser sampling interval is at 40 cm, pixel size of the image is at 8 cm; taken from [Leberl et al., 2010a].

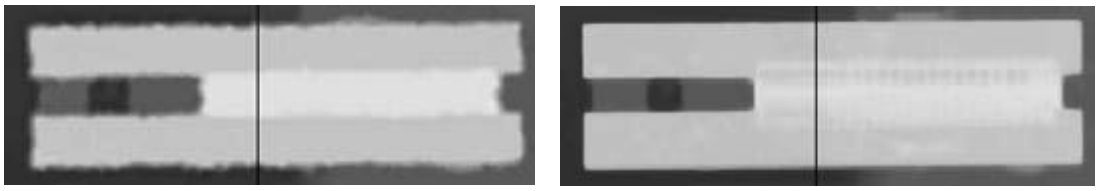


Figure 23: ALS laser scan to the left, photogrammetric 3D to the right. Sampling interval of the ALS at 45cm, of the aerial photography at 8 cm; taken from [Leberl et al., 2010a].

Comparison of Dense Matching Results from StreetSide Imagery and Terrestrial Laserscanning

We have repeated the aerial data analysis also with terrestrial data. Figure 24 is the result of a laser scanning survey of random streets, which were commercially contracted with the British provider StreetMapper [Kremer et al., 2007; Streetmapper]. Figure 25 superimposes over a street-side color photo the laser points of the same area. The photography was collected from a car with high overlaps and with a pixel size of 1.3 cm or 6000 points per m^2. The laser produces 1% of the data from the imagery,namely 60 points per m^2.

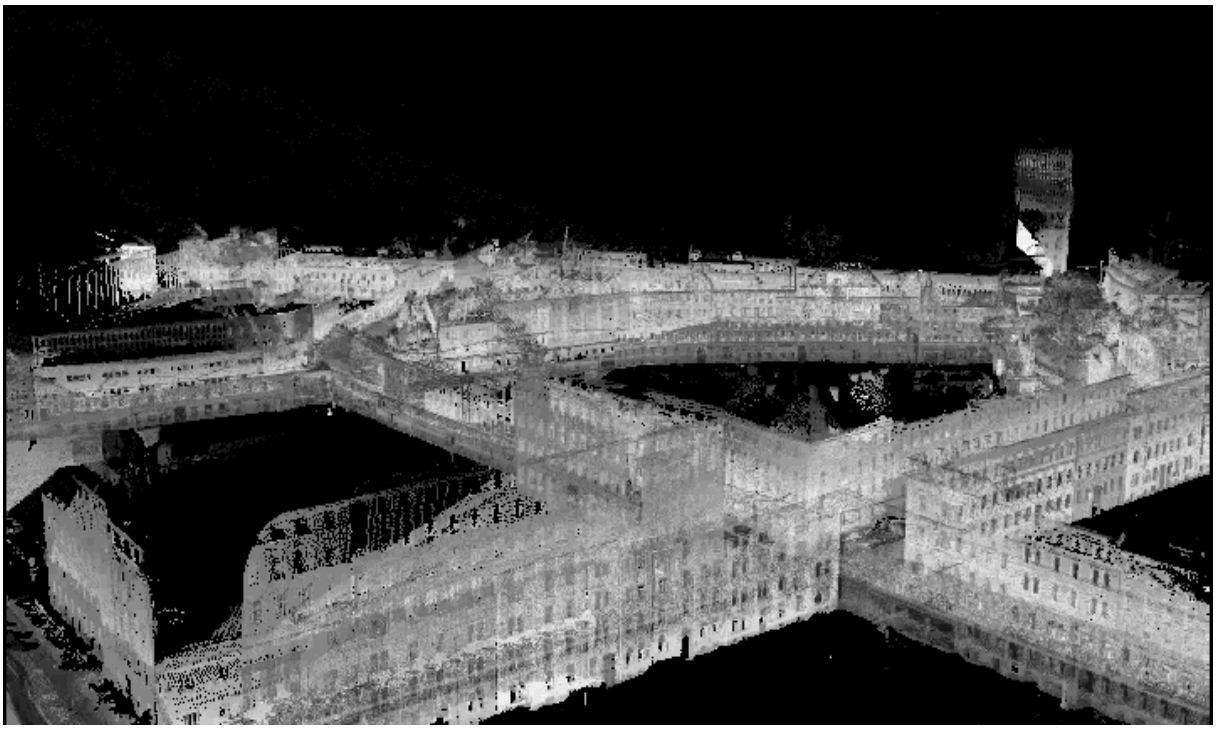


Figure 24: Laser scan of random city streets obtained with the StreetMapper (UK); image taken from [Leberl et al., 2010a].

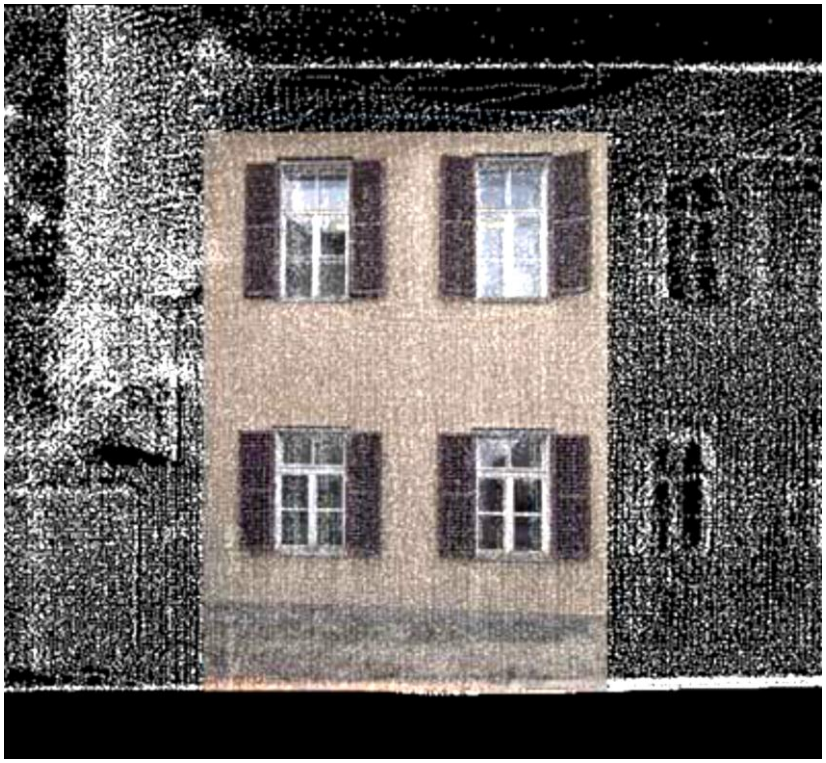


Figure 25: 60 laser points/m2 and 6000 image pixels per m2; taken from [Leberl et al., 2010a].

To assess the photogrammetric 3D point generation workflow, we selected a facade (see Figure 26). We led the imagery through a triangulation without any prior assumptions about the exterior orientation of the cameras, followed by a dense match on a GPU-supported PC via overlapping triplets [Zach et al., 2007b]. Figure 27 illustrates the result of the triangulation with its sparse point cloud, and the arrangement of triplets for the 18 images overlapping a certain facade to produce the dense matches. Figure 28 (left) shows one point cloud patch from one of the 16 triplets, and Figure 28 (right) illustrates the resulting complete point cloud obtained by fusing the 16 patches using a robust fusion method [Zach et al., 2007b]. The approach selected here did not take advantage of the stronger intersection geometry between non-neighboring images. We will investigate those benefits in a separate effort. We have four passes of the laser scanner over the facade shown in Figure 26. This supports an analysis of the differences in depth at ± 1.7 cm among the four passes. Looking at a flat surface patch, we can determine from Table 10 that each laser point cloud has noise at ± 2 cm; this is consistent with the manufacturer's stated specification.

Figure 26: Scene for an automated point cloud from street side imagery with a pixel size of 1.3 cm; taken from [Leberl et al., 2010a].

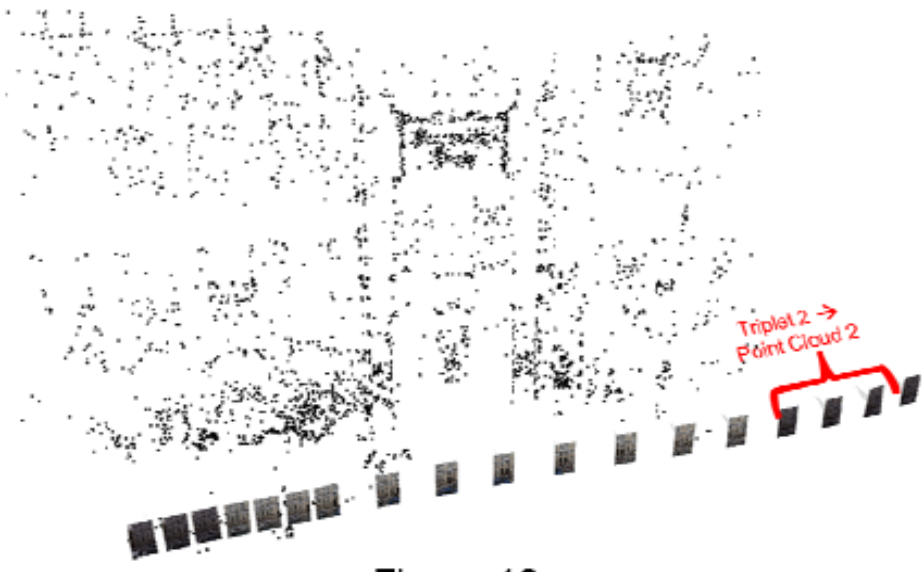


Figure 27: An automated triangulation connects 18 images with overlaps in the selected scene (Figure 26); taken from [Leberl et al., 2010a].

This needs to be related to the photogrammetry result. The error of the mean is represented by an error of the triangulation, typically within a pixel or ± 1.3 cm. The noise of the photogrammetric points is assessed in the same way as the laser's and is found to be at ± 0.9 cm. The experiment supports the conclusion about these two data sets, which is that they are around the same geometric accuracy, with a much higher density for the photogrammetric data.

Figure 28: left: Individual point cloud patch resulting from an image triplet and processed on a PC with one GPU, taking about 2 seconds. The facade is partly obscured by bicycles, moving people and a car; taken from [Leberl et al., 2010a]. right: Fusion result obtained from 16 separately computed patches of point clouds, each from its triplet; taken from [Leberl et al., 2010a].

Laser Points	*Delta [cm]*	*Deviation of the Mean [cm]*
Pass A	2.09	0.66
Pass B	1.28	2.39
Pass C	1.60	1.20
Pass D	1.22	1.84

Delta of all Passes	*Delta [cm]*	*Deviation of the Mean [cm]*
R.M.S.E. of all Means	1.59	1.66

Table 10: Error values of the street-si de laser data in centimeters, taking advantage of the 4 passes of the Krones School. The "Delta-values" are the r.m.s. deviations of the points from a plane in each pass separately since the points are of a vertical wall. The "Deviations of the Mean" are the differences of each point cloud's mean from the overall mean formed from all points in all 4 point clouds. We find a noise level in the 2-cm-range; taken from [Leberl et al., 2010a].

While we have been presenting evidence of a superior photogrammetric process, there is no acceptance of the photogrammetric approach over LiDAR scanning. This is foremost a result of recent years and the failure of photogrammetric technology to innovate. Innovations have been achieved in the sensor domain (the cameras), but not in the analysis domain, thus in the end-to-end automated workflow from images. While some innovations have been achieved, there are three factors that have been preventing their wide-spread acceptance. First, the concept of high image overlaps is objected to with the argument that there is a data overload. To counter that argument one needs to point to the cost of storage that over the recent 3 decennia has eroded at a rate of 12 million; this exceeds the improvement predicted by the rule of [Moore, 1965]. Today a Terabyte disk is available for EUR 100. Second is the concern that photogrammetry needs an aerial triangulation and lasers do not, and that an aerial triangulation consumes resources. However, this concern ignores the fact that the aerial triangulation produces a higher geometric accuracy than the GPS/IMU. Regarding the resources needed for the triangulation, this is a function of the level of automation of the triangulation. The recent availability of a fully automated triangulation in the form of the UltraMap software is meant to alleviate this concern [Hui, 2009].

Third, is the problem often mentioned that dense matching of images is a slow process at a rate of approximately one hour per large format digital image such as those from the UltraCam camera. In the example of a 150-sqkm-city with its 3000 images, the dense match would have taken 3000 hours or 125 days [Leberl et al., 2010a]. Therefore, some photogrammetric practitioners have embarked on so-called blade computers with many CPUs, even hundreds, per system (see for example the Verari blade system www.verari.com). In this context, we point to the illustrations in Figures 11 and 12. The processing of one triplet takes 1 to 3 seconds on a PC augmented by a single graphical processing unit (GPU). We have great expectations that with up to 4 GPUs per PC, one can achieve an acceleration of 400 over an unaided PC. A simple PC with GPU hosting some dense matching software should eliminate the concern over excessive computing times, since with such a system the city-project would be matched in a single 8-hour-shift.

Advantages of Aerial Mapping

Table 11 summarizes the 16 advantages of the photogrammetric approach over the LiDAR-approach. An important factor is that the LiDAR cannot eliminate the need for imagery, because this provides much-appreciated photographic texture.

ACCURACY AND ERRORS
1. Large area rigid camera block geometry via AT
2. Error checking via redundant observations
3. Internal accuracy measures from redundancy
4. Geometric accuracy by AT superior to a reliance on GPS/IMU
5. Greater point density ➔ for better defined discontinuities
ECONOMY
6. Superior data collection efficiency with faster vehicles, larger swaths
7. Single workflow within aerial application, all image based
8. Single workflow across applications (aerial, street side & indoor)
9. No occlusions via no-cost along track high image overlaps
DATA TYPES
10. 2D image information vs. a collection of single 3D data points
11. Multi-spectral image classification
12. Urban facades available at no cost from the air at images edges
13. Images document details ➔ example street signs
MISCELLANOUS
14. Perspective exists towards real time 3D Vision via „*supercomputer in match box*“
15. Full automation needs image redundancy ➔ laser adds little to automation
16. Scene interpretation needs imagery ➔ laser adds little to automation

Table 11: Sixteen advantages of the photogrammetric 3D workflow over the directly measured laser point clouds; taken from [Leberl et al., 2010a].

Therefore, when one obtains the imagery in the appropriate manner, the laser will not add much new value. It would seem that it adds unnecessary cost and complexity to maintain two systems, two sensors, and two workflows when one single system and workflow would suffice. Advocates of LiDAR sensing are working on methods to perform "vision" with the laser point clouds. An example may be the characterization of facade details (windows) from laser point clouds [Haider et al., 2007]. Laser operators add camera pixels as a secondary information source to cope with some of the laser deficiencies. Again, this causes a need for a dual track workflow for both vision and laser. The unique selling proposition (USP) of scanning lasers is the ability to look between leaves on trees onto the ground and thus provide multiple reflections of the laser so that a bald Earth terrain model as well as a surface model can be produced with ease. To

achieve the same split from photography into a bald Earth and its vertical objects one will need significant automated processing for object classification and detection of vertical objects, as well as some manual post-editing to refine the results. A meaningful comparison between the separation results from both technologies will have to be subject to future studies.

Advantages of Street Side Imagery

The appetite for detail justifies an augmentation of aerial mapping by street-level and indoor mapping. Commercial solutions have been developed in significant numbers for LiDAR to map streets (Figure 24), and indoor spaces such as tunnels [Kremer et al., 2007; Streetmapper; Leica, 2008; Zoller + Fröhlich, 2007], as well as for camera systems [Cyclomedia, 2008; Facet, 2008]. The workflow and technology issues also transfer into these newer applications. The comparison between street-side and indoor lasers versus the use of blocks of overlapping street-level and indoor digital images is of great interest, but has not been widely studied. One effort was by [Strecha et al., 2008] using LIDAR data [Zoller + Fröhlich, 2007] as ground truth to assess computer vision results. The authors show that current fully automated Structure-from-Motion pipelines like [Martinec et al., 2007] achieve camera calibration accuracy close to the 3σ value of the LIDAR ground truth. Our contribution in this paper adds to this quantitative accuracy evaluation.

Besides the geometric accuracy and the concerns for the economy of data acquisition, there exist various qualitative issues to consider in a comparison:

- Maintaining single workflows in lieu of separate workflows will be commercially attractive, particularly if the uniform workflows are applicable across various mapping scenarios, be they aerial, in the streets or indoors.
- Images document the details of street sides and indoor scenes, for example with street signs and alpha numeric information on facades.
- The interpretability of objects from imagery as a computer vision task relies on imagery. Therefore, both images and redundancy are unavoidable. The laser may not add much value to a redundant block of street side or indoor images when an automated workflow for 3D vision is in use to build models of all objects for subsequent selective visualization.

Conclusions

[Grün, 2009] has argued that statements about the demise of photogrammetry are premature. We do support that vision by a specific look at a major factor in the presumed demise, the scanning LiDAR-technologies, and we compare their results with 3D points from photogrammetry. We are investigating the geometric accuracy and point densities from airborne and street-side LiDAR scanners versus highly overlapping imagery and fully automated photogrammetry. While it seems fully understandable that LiDAR scanning will be advantageous over film-based 3D photogrammetry, that advantage goes away as photogrammetry morphs into 3D vision based on a fully digital and automated workflow. We can show that in aerial situations, the accuracy of 3D-photogrammetric points is often better than ± 1 pixel, whereas the LiDAR accuracies are defined by the direct geo-positions from GPS/IMU measurements. Aerial point densities favor the image- derived data over the direct LiDAR observations. In the application to street-scenes, the data densities differ more strongly. The image-based data may be 100 times richer than the LiDAR points while the geometric accuracies seem to be in the

same range with values at ± 1 to 2 cm. Apart from the geometric parameters of these technologies, there are numerous economic considerations as well as considerations concerning the richness of the information content; these all favor the image-based approach. With an aerial camera, a mission may be completed in 10% of the time for a LiDAR-mission. In street-side applications, the ability to identify signs, to characterize the scenes, to track their changes and to be much less dependent on directly measuring the sensor pose may be an invaluable advantage over the simple collection of point clouds with a direct pose measurement.

4.13. Experiments: Oblique versus Vertical Photography

This sub chapter is reprinted from [3]. Oblique aerial photography has become a widely used resource for urban imaging. Originating in the US and championed by Pictometry, oblique images are now being acquired world-wide. We are interested in a comparison between oblique and vertical aerial photography, especially addressing the facade in urban areas and facade details such as the number of floors and windows. In this section, we want to compare the geometric resolutions of both imaging technologies with a special interest in facade details. Additionally, we will clarify the question of which technology is better suited to characterize building facades, and in chapter 7 we will go into detail concerning the results of the characterization of facade details.

Figure 29: Two aerial images showing Herz Jesu Church in Graz; (left) oblique aerial image; Detail from Microsoft Bing Maps. (right) vertical aerial image

Oblique Camera Geometry

The oblique images, we use for our comparison, have been produced with a Maltese Cross configuration. While exact technical data of such oblique images remain unavailable, we have taken it upon ourselves to reverse engineer them.

The oblique images have 2,672 rows and 4,000 columns representing a format defined from traditional consumer cameras at 36 mm x 24 mm like the oblique camera MIDAS built by Track 'Air. This results in a pixel pitch of 9 micrometers.

Well-mapped terrain with large vertical structures including a church (see Figure 29) can serve as a test area to compute a resection in space with self-calibration, in which we compute the internal geometry with its focal length. Taken together with the focal plane format, this results in an angle of view of 24 degrees x 16 degrees. The angle of view from near range to far range varies by 16 degrees if the CCD array is arranged with the rows perpendicular to the "range direction". Let us take a look at the 4 corners of the image, as well as the camera station, in lat-long geographic coordinates (Table 12):

Decimal Degrees	Lat	Long
UpperLeft	42.39402019	-83.09217033
UpperRight	42.39427638	-83.08296264
LowerRight	42.38810587	-83.08379572
LowerLeft	42.3878802	-83.09028031
Centroid	42.39107066	-83.08730225
Camera Station	42.38016659	-83.08618271
Relative to Lower Left	Lat	Lon
LowerLeft	0	0
LowerRight	0.00022567	0.00648459
UpperRight	0.00639618	0.00731767
UpperLeft	0.00613999	-0.00189002
Centroid	0.00319046	0.00297806
Camera Station	-0.00771361	0.0040976
1 degree being m	111,000	82,206
LowerLeft	0	0
LowerRight	25	533
UpperRight	710	602
UpperLeft	682	(155)
Centroid	354	245
Camera Station	(856)	337

Table 12: Image parameters derived from oblique aerial images extracted from BingMaps website [from Prof. F. Leberl]

This indicates that from a flying height of 1130 meters, the near range is about 850 m away, while the far range is at 1530 meters. The foot print is shown in Figure 30. These results coincide with the results that are provided by the company Vision-me. They quote a front line distance of 550m, a back line distance of 770m and a side line distance of 700m.[20]

[20] [http://vision-me.com/Factsheet%20Pictometry.pdf]

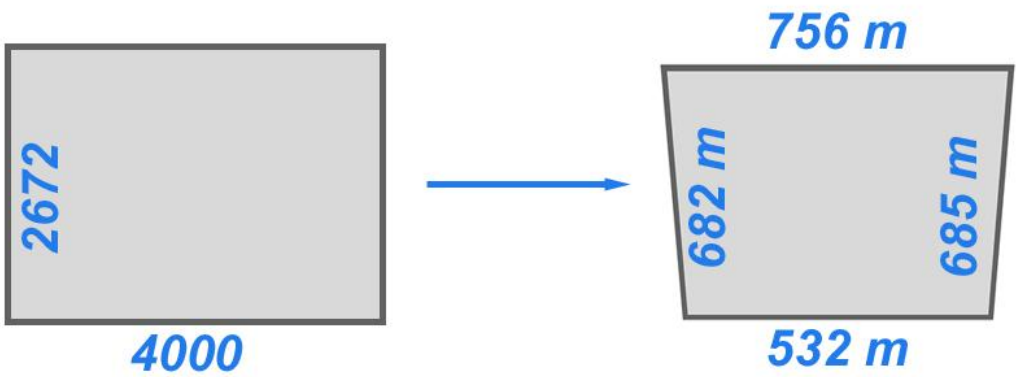


Figure 30: Projecting a rectangular image format onto a trapezoidal shape on the ground. The camera position is 850 off from the near range [from Prof. F. Leberl]

This leads to a near range off-nadir angle of 37 degrees and a far range off-nadir angle of 53 degrees. The most important results are summarized in Table 13. The along track GSD values resulting from those angles and the camera parameters with the 9 micrometer CCD elements and a focal length of 85 mm are as follows:

Near range…… ~ 14 cm (4000 pixels across a distance of 530 m)

Far range…… ~ 19 cm (4000 pixels across 760 m)

This is also the geometric resolution not in the ground plane, but in vertical planes such as the facades represent, but in the "range-direction". In the "azimuth-direction", the geometric resolution is less, since 680 m are mapped into 2672 pixels, resulting in 25 cm per pixel.

Pixel size in the image plane =	9µm
Focal length =	85.5mm
Viewing angle of the camera =	24° x 16°
Flying height above ground =	1,130m
Distance to near range =	850 m

Near range off-nadir angle =	37 °
Far range off-nadir angle =	53 °
Horizontal GSD at near range =	14 cm
Horizontal GSD at far range =	19 cm
Distance to far range =	1530 m

Table 13: BING/Maps oblique imagery parameters reconstructed from known terrain points.

Vertical Image Geometry

Table 14summarizes some relevant geometric parameters of a digital aerial camera in the form of the UltraCamX and wide angle UltraCamXP.

	UltraCamX	*UltraCamXP - WA*
Image Rows x Columns	14, 430 x 9,420	17, 310 x 11,310
Image size in X and Y, [mm]	103.9 x 67.8	
Pixel size in image plane [µm]	7.2	6
Focal length[mm]	100	70
Max Look angle off-nadir [deg]	27.5	36.5

Table 14: Some geometric data of two typical digital aerial cameras (from www.vexcel.com)

Visual Comparison

One can image facades in vertical aerial imagery at the image's edge under an angle of 27 °. With new wide angle systems, this angle increases to 35 °. Oblique cameras produce larger angles at 35 to 55 °. Because the viewing angles of facades are typically at 45° and thus larger in oblique than in vertical photography with its 27° at the image's edges, one might be tempted to consider oblique sources in the façade analysis.

For this purpose we can access high quality vertical aerial photography from commercial sources taken, in our case, with the UltraCam-series of aerial cameras and oblique aerial imagery coming from theMicrosoft's BING/Maps mapping website. Figure 31 presents an example of an area in Graz (Austria) with a vertical coverage and superimposed the outlines of an oblique image. Also shown is the oblique image itself in its original geometry.

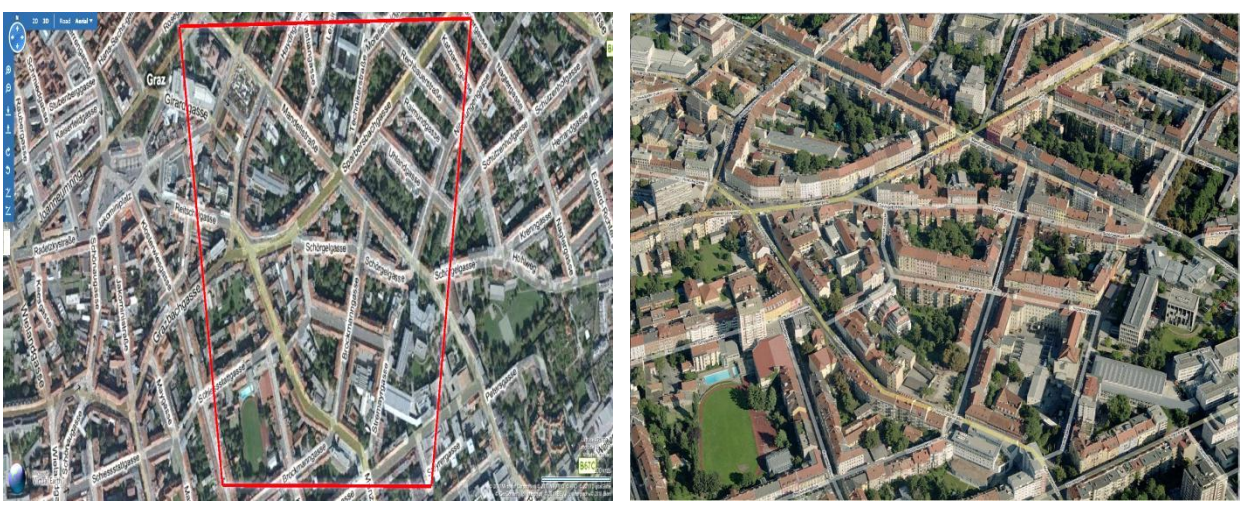


Figure 31: Detail from Microsoft Bing Maps. Left is the orthophoto and superimposed outline of an oblique aerial image produced with the Pictometry system operated by Blom. Right is the oblique aerial image.

A visual comparison of vertical versus oblique images in Figure 31 and Figure 32 does not result in a clear advantage of one versus the other approach. At an off-nadir angle of about 45°, the oblique images have more significant occlusions, given that the vertical images show the same facades at an angle of only 27°. Regarding the radiometric range, one would give the vertical

images an advantage. This visual impression from original images is overwhelmed by the differences in geometry. To eliminate that factor, we create rectified versions in the plane of the facade, as shown in Figure 33. Again, a visual comparison does not show a clear advantage of one technology over the other. For this reason, in the next chapter, we have to determine the pixel sizes on building facades for both imaging techniques, which will provide us with more reliable statements about the geometric accuracies.

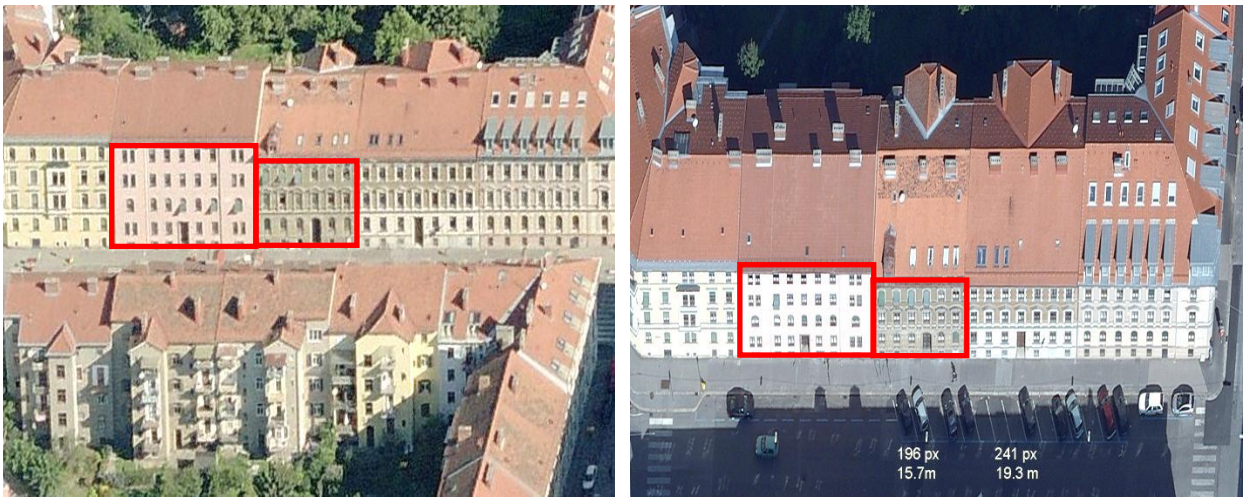


Figure 32: Oblique aerial image at 45° look angle taken from Microsoft Bing Maps (left); Vertical aerial image obtained from UltraCamX at a look angle of 27° (right). The marked sections in the original image segments (above) have been rectified (below).

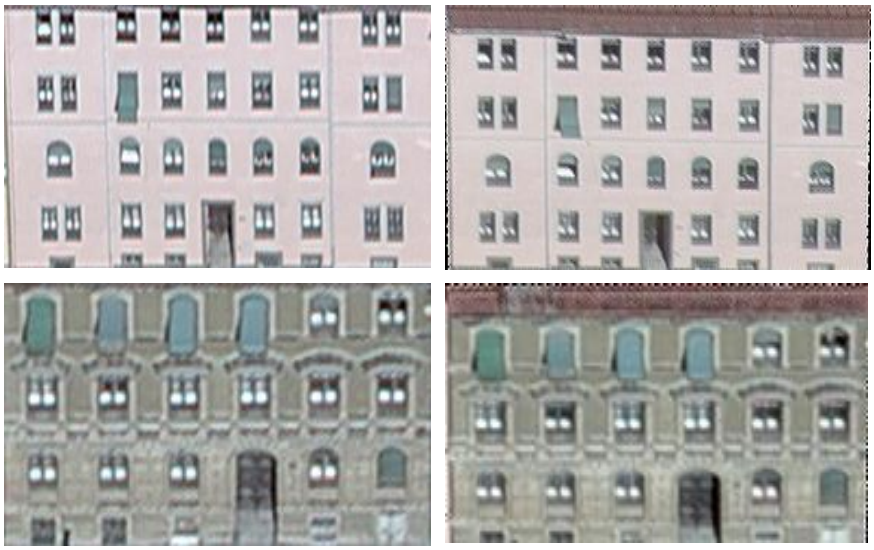


Figure 33: The marked sections of Figure 32 have been rectified. At left are two sections from the oblique data at 45°, at right from the vertical data at 27°.

Pixel Sizes on Façades

For a vertically-looking camera, the pixel on a facade (FSD or Facade Sampling Distance) changes as a function of the look angle off-nadir a with

$$FSD = \frac{GSD}{\tan(\alpha)}$$

These results in Table 15 for a GSD at 10 cm, a typical value for urban aerial photography. The facade pixels are rectangular.

Angle (deg)	0	5	10	15	20	25	30
Pixel vertical [cm]	∞	114	57	37	27	21	17

Table 15: Incidence or look angles and vertical pixel size within a facade. The horizontal pixel size is at 10 cm.

For an oblique camera, the pixel size within a vertical plane is defined by two angles (see Figure 34). Angle β is the orientation of the optical axis off-nadir. Angle a is the angle between the optical axis and the actual imaging ray. The off-nadir angle β produces two GSD values, one in the direction of the inclination of the optical axis GSD_r, with r being the range direction or direction between nadir and the optical axis; and the GSD_a in azimuth direction

$$GSD_r = \frac{p * H * \cos(\alpha)}{(f * \cos^2(\beta))}$$

$$GSD_a = \frac{p * H * \cos(\alpha)}{(f * \cos(\beta))}$$

A vertical facade is resolved as a function of where in the image it is located and this is defined by the second angle a, producing a vertical pixel dimension FSD_v:

$$FSD_v = \frac{p * H * \cos(\alpha)}{\left(\frac{f}{\sin(\beta)}\right)}$$

Table 16 presents some vertical facade pixel sizes for the oblique camera with a look angle at 53° and compares this to pixel sizes from an UltraCamX with an angle at 22°. In this example, the vertical facade pixel sizes for the oblique camera at an angle of 53° and for an UltraCamX at an angle of 22° are almost identical. The simple conclusion is that the pixel size not only is a function of the look angle, but also of the flying height and the GSD. While the look angle appears a lot less attractive from a consideration of look angles in the UltraCam, on a given facade, this does not propagate into an inferior geometric resolution.

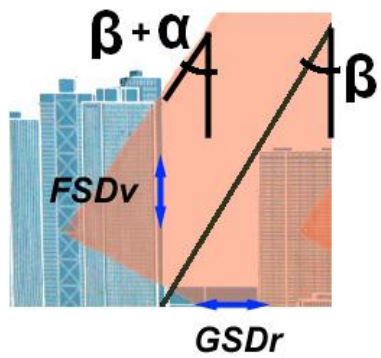


Figure 34: Side looking camera with off nadir angles and GSD in "range direction" (GSDr) and pixel size on the facade (FSDv).

	Degrees	*Azimuth (cm)*	*Range (cm)*	*Façade vertical (cm)*
Oblique camera (near range)	37	14.7	20.9	27.7
Oblique camera (far range)	53	19.6	27.7	20.9
UltraCam-X	22	8.1	8.1	20.0

Table 16: Size of a pixel on a facade in cm, as a function of the look angle, in degrees

Efficiency of Aerial Data Collection

A consideration of image pixel sizes ignores the efficiency of one versus another imaging approach and technology. Flying at a certain height to achieve small pixels, and producing images with a large format, will be more efficient than flying with small formats for a small swath width and at a low flying height. An UltraCam, for example, produces 17.5 K pixels in one single flight line. An oblique camera will have to match this number to be comparatively productive. At a frame size of 4,000 pixels, one will not easily match the productivity of a vertical mapping camera.

Conclusions

The efficiency of aerial imaging may favor vertical technologies over the oblique approach. Vertical images today produce 280 Megapixels per exposure, whereas oblique cameras still operate at the 10-50 Megapixel level. Even if one were to consider that in a Maltese Cross arrangement one operates with 5 such cameras, this still adds up to only 50 Megapixels.

A limitation of the current "normal angle" aerial cameras is the look angles one can achieve in the direction of flight at perhaps 17° off nadir. Solutions are either a cross flight pattern, or the use of a wide angle camera model such as the UltraCamXp-WA with 26° in flight direction, or the use of the new single CCD-chip DMC-II digital aerial camera by Intergraph.

Going beyond a mere "eye candy" approach for the use of oblique images, one will quickly find that novel high-redundancy vertical aerial images offer a superior source of information about urban areas, street canyons and facades. We suggest that the benefits from vertical

aerial photography have been undervalued, and that conversely benefits from oblique images have been overstated.

Another big drawback of oblique cameras is the lower radiometric resolution. In contrast, digital aerial cameras produce images in the four colors: red, green, blue and near infrared (NIR), which enables more complex characterizations of urban structures.

4.14. Discussion and Summary

All image-based source data for the proposed property work are the result of extensive computation and data processing, some of it constituting the outcome of considerable recent innovations, such as dense image matching and fully automated aerial triangulation. However, none of that processing is specific to the property characterization; therefore, we do not address it in this paper.

The growing diversity of Geodata sources has recently been discussed by [Blumenthal, 2009] with respect to efforts made at Google. Two innovative phenomena stand out: crowd sourced imagery and motion traces. Both may have an impact on the property characterization as such data becomes increasingly more available. The detail from crowd sourced imagery (also denoted as Community Photo Collections CPC) will most likely exceed by a level of magnitude. In 2011, more than 60 billion pictures[21] are taken by people worldwide and more than 100 million photos are uploaded to the web every day[22]. The developments in this domain hold great promise [Agarwal et al., 2009; Gösele et al., 2010; Snavely et al., 2008]. Even new words are being introduced to deal with these changes, for example "neo-geographer" [Goodchild, 2008] and neo-photogrammetry. [Leberl, 2010c]

[Leberl, 2010c] describes neo-photogrammetry as follows: "*Considering the need to use street-level imagery and images of the interiors of important buildings or shopping malls, one will quickly expect that such images will come from us, the universe of Internet users. We are already feeding images at a rate of 1 million per day into FLICKR and similar sites. We envision a geo-data-future dominated by community- contributed images and Internet-resident automated photogrammetry solutions. One may well find an entirely new mapping paradigm, away from industrial solutions and defined by the global community of Internet-users, each one a potential neo-photogrammetrist.*"

[21] http://www.itfacts.biz/50-bln-digital-photos-taken-in-2007-60-bln-by-2011/8985
[22] http://royal.pingdom.com/2011/01/12/internet-2010-in-numbers

CHAPTER FIVE

DATA PER PROPERTY

Contents

5.1	Introduction	107
5.2	Coordinate Transformations	108
5.3	Image Matching	113
5.4	Minimizing Building Overhangs and Mismatches	119
5.5	Building Façade Separation	120
5.6	Data per Property	125
5.7	Experiments	125

5.1. Introduction

After we have generated all necessary input data, we can now use this data for the real estate characterization. A central task exists to identify "buildings". The definition of a "building" is less obvious than it may initially seem. The imagery needs to be related to parcel maps in the form of cadastral records. Figure 45 presents a cadastral map segment and superimposes it over the orthophoto. The first observation concerns the geometric relationship: the visual data from the imagery are not in complete agreement with the cadastral parcels, and a geometric change is needed to achieve an optimum match. The second observation concerns the fact that buildings, as seen in aerial imagery, cut across property boundaries because they may be attached to one another in dense urban situations.

What then is a "building"? In our context, this is a structure of sufficient size within a parcel. Therefore, what may be experienced as a single building in aerial photography will be represented by a collection of buildings, each defined by its own parcel. The inverse may also exist, where multiple buildings are defined on a single parcel. This fact leads to a third issue, namely a need to separate smaller structures such as garages or sheds from a building proper. A fourth topic addresses complex building shapes with many facades. For analysis purposes it would be desirable to have buildings with only 4 facades. An approach to cope with the complex building shapes may consist of separating an individual building into its parts so that fairly basic building shapes are then achieved, in analogy to separating the concatenated buildings of urban landscapes along parcel boundaries.

To be able to characterize buildings that are situated on these parcels we have to split a visible whole building block into its different single buildings. Usually this is done using cadastral information to divide the single land parcels. Most cadastral maps, as well as the Austrian cadaster, basically present a 2D data base and ignore the 3rd dimension. This causes issues when relating the cadastral data to the aerial photography and its inherently 3D data products. Therefore, co-registration may be accomplished using the 2D orthophoto and a match between it and the 2D-cadastral map. Once this co-registration is achieved, the cadastral data are also geometrically aligned with all the DSM and other photo-derived data sets. A 3-step process serves to match the 2D cadastral vector map (in its own coordinate reference system) and the 2D orthophoto (in its different coordinate reference):

1. Coordinate Transformations
2. Image Matching
3. Minimizing Building Overhangs and Mismatches

But why are these three processing steps necessary? In the next section, we will clarify this question. Another question we would like to answer is: "What can we do if the cadastral boundaries don't divide two buildings in a correct way or if several attached buildings belong to a single cadastral parcel? In both cases, we need a method to correctly split these buildings. But how can we split buildings, if the cadastral information is not useful?" We present a facade separation method that is able to deal with such cases.

5.2. Coordinate Transformations

Coordinate Systems and their transformation are a crucial topic when characterizing vertical aerial images. Especially in the field of computer science coordinate systems and their transformations are sometimes handled with less care, which leads to nice visualizations but inaccurate results. Therefore and for the ease of reference we will review some general aspects.

A coordinate reference system defines how geo-spatial objects relate to real locations on the Earth's surface. It combines a coordinate system with a datum, which gives the relationship of the coordinate system to the surface and shape of the Earth. A coordinate system may also include a coordinate transformation (traditionally called a map projection) that relates positions in geodetic coordinates to positions on a 2D map or a 3D virtual Earth.

Datum

To be able to georeference objects on the Earth's surface we have to define how the surface is described. The parameter or the set of parameters that serve to describe the surface are called datum. It defines the position of the origin, the scale and the orientation of the coordinate system. There are a horizontal and a vertical datum .

Horizontal Datum

Because of Earth's rotation, which creates a centrifugal force perpendicular to the rotation axis, the Earth is not a perfect sphere but rather an approximated ellipsoid that is flattened at the poles. . If the Earth consisted of solid material, then there would be no effect on the shape.

However, our Earth has a molten interior, it has tectonic plates on the thin crust. The Earth is "viscous" and this accounts for this slight flattening at the poles.

In geodetic surveying, the computation of the geodetic coordinates of points is commonly performed on a reference ellipsoid closely approximating the size and shape of the Earth in the area of the survey. This is also called the horizontal datum because it is the model used to measure positions on the Earth. It is defined by the equatorial radius (semi-major axis) and the flattening at the poles (semi-minor axis). This flattening is calculated by determining the quotient of the difference of the equatorial and the polar radius and the equatorial radius. Table 17 lists the most common reference ellipsoids:

Name	Equatorial axis [m]	Polar axis [m]	Inverse flattening [1/f]
Airy 1830	**6 377 563.4**	**6 356 256.9**	**299.324 975 3**
Clarke 1866	**6 378 206.4**	**6 356 583.8**	**294.978 698 2**
Bessel 1841	**6 377 397.155**	**6 356 078.965**	**299.152 843 4**
International 1924	**6 378 388**	**6 356 911.9**	**297**
Krasovsky 1940	**6 378 245**	**6 356 863**	**298.299 738 1**
GRS 1980	**6 378 137**	**6 356 752.3141**	**298.257 222 101**
WGS 1984	**6 378 137**	**6 356 752.3142**	**298.257 223 563**

Table 17: Datum for the most common reference ellipsoids
[http://en.wikipedia.org/wiki/Reference_ellipsoid]

The reference ellipsoid calculated by Friedrich Bessel in 1841 is the basis for the geodetic surveys in Austria, Germany and Switzerland. The Hayford ellipsoid was named by the US geodesist John Fillmore Hayford, and was also referred to as the International ellipsoid in 1924. The Krassovsky ellipsoid is the reference for all geodetic surveys in Russia and former Eastern European countries. Another very important reference ellipsoid is the WGS 84 that builds the basis for Global Positioning Services (GPS). It is compared to the other mentioned reference ellipsoids as a global solution that approximates the whole Earth and the close space as close as possible.

Vertical Datum

Height measurements on the other hand are performed based on the so called vertical datum. In common usage, elevations are often cited in height above mean sea level (MSL). The MSL is defined as the arithmetic mean of the hourly water elevation taken over a specific 19 year cycle. This definition eliminates tidal highs and lows due to the gravitational effects of the sun and the moon. The MSL is defined as the zero elevation for a local area. However, zero elevation as defined by one country is not the same as zero elevation defined by another, because the sea levels are different depending on the geoid.

This means that all height measurements that are made on the surface of the Earth are referred to the geoid. "*The geoid coincides with that surface to which the oceans would conform over the entire Earth if free to adjust to the combined effect of the Earth's mass attraction (gravitation) and the centrifugal force of the Earth's rotation. As a result of the uneven distribution of the Earth's mass, the geodial surface is irregular and, since the ellipsoid is a regular*

surface, the separations between the two, referred to as geoid undulations, geoid heights, or geoid separations, will be irregular as well."[23]

Although the physical Earth has excursions of +8,000 m (Mount Everest) and −11,000 m (Mariana Trench), the geoid's total variation is less than 200 m (−106 to +85 m) compared to a perfect mathematical ellipsoid.[24]

Coordinate Systems

After we have defined the relationship of the coordinate system to the surface and shape of the Earth, we will discuss the basics of coordinate systems. A coordinate system uses a set of numbers, or coordinates, to uniquely determine the position of a point in a multi-dimensional space. In geodesy, there are two main coordinate systems [Bretterbauer et al., 2003]:

- Geographic Coordinate System
- Cartesian Coordinate System

Geographic Coordinate System

With geographic coordinates it is possible to determine the exact position of a point on the Earth's surface. Therefore, the Earth is divided into 360 lines of longitude and 180 lines of latitude. The geographic latitude (abbreviation: Lat., φ, or phi) of a point on the Earth's surface is the angle between the equatorial plane and a line that passes through that point and is normal to the surface of a reference ellipsoid, which approximates the shape of the Earth. The longitude (abbreviation: Long., λ, or lambda) of a point on the Earth's surface is the angle east or west from a reference meridian to another meridian that passes through that point. All meridians are halves of great ellipses, which converge at the north and south poles.[25] The third parameter to determine the exact position of a point is the height.

Cartesian Coordinate System

Besides geographic coordinates in geodesy, Cartesian coordinates are also based on a map projection. The big difference to Cartesian coordinate systems in mathematics is that the x and y-axes are exchanged. The origin of such coordinate systems is defined arbitrarily depending on the dimensions of the respective area to optimize the representation.

Another very important aspect is the map projection, which is crucial to model the surface of the Earth on a plane. Therefore, more generally, a map projection is any method of "flattening" into a plane a continuous surface having curvature in all three spatial dimensions. Carl Friedrich Gauss proved in his "Theorema Egregium" that a sphere cannot be represented on a plane without having distortions. Thus, the goal of every map projection is it to minimize these distortions and to preserve one or two (but never all) of the following properties:

- Equidistant (distance preserving maps)
- Equal-Area (area preserving maps)
- Conformal (angle preserving maps)

[23] [http://en.wikipedia.org/wiki/Figure_of_the_Earth]
[24] [http://www.csr.utexas.edu/grace/gravity/gravity_definition.html]
[25] [http://en.wikipedia.org/wiki/Geographic_coordinate_system]

Due to the many uses of maps and the vast range of scales at which they are created, no single map projection serves well for all purposes. The projections are described in terms of placing a surface in contact with the Earth. These projection surfaces can be cylindrical, conic or plane (azimuthal). Once a choice is made between projecting onto a cylinder, cone, or plane, the aspect of the shape must be specified. The aspect describes how the developable surface is placed relative to the globe: it may be normal (such that the surface's axis of symmetry coincides with the Earth's axis), transverse (at right angles to the Earth's axis) or oblique (any angle in between). The developable surface may also be either tangent or secant to the sphere or ellipsoid. Tangent means the surface touches but does not slice through the globe; secant means the surface does slice through the globe.

Azimuthal projections touch the Earth's surface in one point. They have the property that directions form a central point are preserved, hence great circles though the central point are represented by straight lines on a map. There exist two different azimuthal projections.[26]The first type is the true perspective projection that maps the Earth's surface by extending lines from a point of perspective onto a plane:

- *Stereographic Projection*: It's a map projection obtained by projecting a point X on the surface of sphere from the sphere's north pole N to a point Y in a plane tangent to the south pole S. Can display nearly the entire sphere on a finite circle. The full sphere requires an infinite map.
- *Gnomonic projection*: displays great circles as straight lines. The gnomonic projection is a nonconformal map projection obtained by projecting a point X on the surface of sphere from a sphere's center C to point Y in a plane that is tangent to a point T.
- *Orthographic Projection*: maps each point on the Earth to the closest point on the plane. The orthographic projection is a projection from infinity that preserves neither area nor angle.

The second type of azimuthal projections is the not true projective projection like:

- *Azimuthal equidistant*: A straight line from the central point on the map to any other point gives the shortest distance between the two points. Azimuthal equidistant maps are used mainly for plotting direction and distance from the map's central point, but measurements originating from other points on the map can be greatly distorted.
- *Lambert azimuthal equal-area*: It is a particular mapping from a sphere to a disk. It accurately represents area in all regions of the sphere, but it does not accurately represent angles.
- *Logarithmic azimuthal:* This type of projection can be used to enlarge a circular region of almost any size, with the surrounding region subordinated to an adjustable extent. It resembles a "fish-eye" photograph

[26] [http://en.wikipedia.org/wiki/Map_projection]

Conic projections project the Earth's surface on a cone. This map projection superimposes a cone over the sphere of the Earth, with one or two reference secants to the globe that intersect it.

Cylindrical projections project circles of longitude to equally spaced vertical lines and circles of latitude to horizontal lines. This mapping of the meridians leads to a stretching of distances in east-west direction on the globe. Depending on the stretching factor there exist multiple cylindrical projections like the Mercator or the Lambert projection.

Using this knowledge we can perform the coordinate transformation. Therefore, the cadastral point coordinates get converted from their coordinate reference system in the reference of the orthophoto. For our test area we deal with the Gauss-Krueger M34 coordinate system concerning the cadastral information and the Universal Transverse Mercator UTM-System.

Gauss Krueger M34 coordinate system

The Gauss-Krueger coordinate system was named after Carl Friedrich Gauss and Johann Heinrich Louis Krueger. It is mainly used in German speaking countries for almost all official topographic maps including cadastral maps. It was introduced to Austria in 1921 as a uniform reference system for topographic and cadastral surveys. It is based on the Bessel ellipsoid with its fundamental point Hermannskogel and a transversal cylindrical projection. It is equidistant at the reference meridian. It describes the Earth's surface using a series of numbered zones that are 3 degrees of longitude in width. The zero meridian is located at the most western point of Europe, the Island El Hierro (former Ferro), 17°40' west of Greenwich (Great Britain). This covers the whole national territory by only using three zones (28°, 31° and 34° east of Ferro).

Universal Transverse Mercator UTM coordinate system

The Transverse Mercator (UTM) coordinate system is very similar to the Gauss Krueger System. There are two main differences: The UTM system divides the surface of Earth between 80°S and 84°N latitude into 60 zones, each 6° of longitude in width and divided by the equator into north and south portions. Each UTM zone has its own planar coordinate system, and each zone has an entry in the Universal Transverse Mercator subgroup within the Predefined list's Global and Regional group (e.g. UTM zone 33N).

The second main difference is that because of the bigger zones a scale factor is used to reduce distortions at the borders of the single zones. In each zone, the scale factor of 0.9996 along the central meridian reduces the diameter of the transverse cylinder to produce a secant projection with two vertical lines of true scale, located approximately 180 km on either side of, and approximately parallel to, the central meridian. The scale is less than 1 inside these lines and greater than 1 outside of these lines, but the overall distortion of scale inside the entire zone is minimized.

Linear conformal coordinate Transformation

For the transformation between the GPS coordinates of the projection centers of the aerial images and the cadastral points we use a 7-parameter Helmert transformation. The used parameters are a translation vector C, a scale factor μ and a orthogonal rotation matrix R consisting of three axes r_X ,r_Y and r_Z

$$\boldsymbol{X}_T = C + \mu R \boldsymbol{x}$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_T = \begin{bmatrix} c_X \\ c_Y \\ c_Z \end{bmatrix} + (1 + s * 10^{-6}) * \begin{bmatrix} 1 & -r_Z & r_Y \\ r_Z & 1 & -r_X \\ -r_Y & r_X & 1 \end{bmatrix} * \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

where X_T is the transformed vector and x is the initial vector. The parameter values used for this coordinate transformation are listed in Table 18.

Region	*Start Datum*	*Target Datum*	*cX*	*cY*	*cZ*	*S [ppm]*	*rX [arcsec]*	*rY [arcsec]*	*rZ [arcsec]*
Austria	WGS84	MGI	- 577.326	-90.129	-463.920	-2.423	5.137	1.474	5.297

Table 18: Standard parameters for coordinate transformation WGS84 – GK (Source: BEV 2011)

In an ideal world, this would solve the registration problem. It does not, as illustrated in Figure 45(left). Thus, a second step is needed to achieve a fine alignment of the cadastre and the orthophoto. This is needed due to small projection errors and errors because of inaccurate GPS/IMU measurements in the range of a few pixels (some tens of cm).

5.3. Image Matching

To reduce these small errors we have to compare the two overlaid images and detect similarities. In general, this is done by comparing the single pixel values x_{ij} and y_{ij} of two images I_1 and I_2. If these values are identical the difference between these values is zero. However, the direct difference is in general not meaningful because there always exist differences due to different lightening and capturing conditions. There exist three classes of matching methods.

In our case, we deal with a true orthophoto with raster information and cadastral boundaries containing vector data. Moreover, the true orthophoto is a RGB image and the cadastral information is a binary image. How can we compare these two data types?

Intensity Based Matching

The first are the intensity based methods that directly use the gray values of an image like cross correlation or Least Square Matching (LSM). These methods deliver results with sub-pixel accuracy; however, they also need good approximated values. These methods use template matching where you are looking for a specific object in a target image. What we want to do is to detect a reference image R(l,j) in a target image T(r,s). The goal is to determine the position in T(r,s) where there exists an optimum correlation between the two visual contents. For this a common practice is to use the Zero mean Normalized Cross-Correlation function (ZNCC):

$$\rho(x,y) = \frac{\sum_{u=-N}^{N}\sum_{v=-M}^{M}\left(f(x'+u,y'+v) - \bar{f}\right) * (g(x+u,y+v) - \bar{g})}{\sqrt{\sum_{u=-N}^{N}\sum_{v=-M}^{M}\left(f(x'+u,y'+v) - \bar{f}\right) * \sum_{u=-N}^{N}\sum_{v=-M}^{M}(g(x+u,y+v) - \bar{g})}}$$

with a template dimension x = 2N + 1 and y = 2M + 1, where f, g are the gray values of the template and the patch, $\bar{f}, \bar{g}$ are the corresponding average values, x', y' and x,y are the pixel coordinates of the centers of the template respectively the patches. The result of this calculation is the optimal position of the template in the target image. This works well if you have two images with similar content like two aerial images with the same radiometric resolution.

Relational Methods

Relational matching is a method for finding the best correspondences between structural descriptions. It is widely used in computer vision for problems like object recognition and location, scene analysis and navigation. Relational description consists not only of features, but also geometrical and topological relationships among the features. The matching algorithm then has to determine which image elements and object model parts correspond. The measure of similarity between two relational descriptions can be achieved by an evaluation function that is usually defined as a cost function. This function has to be minimized and is zero if two descriptions are identical.

Feature Based Matching

The feature based matching consists of two main processing steps: 1) the extraction of features like edges or points in an image and 2) the matching of these features within the images.

To be able to handle these two data sources we have to use this method. Because we deal with two different image formats we have to extract edges in these images using two different ways. Concerning the true Orthophoto we have to detect edges in the grayscale image. Concerning the cadastral information consisting of vectors, a line rasterization of the vectors is necessary. Thus, the following processing steps are necessary and are described in more detail:

- Edge Detection
- Line Rasterization
- Feature Matching

Edge Detection

Edges are properties of pixel pairs, in which the function changes abruptly. The advantage of edge detection is that it leads to a significant reduction of image data that is crucial when dealing with large data amounts. There exist a huge collection of different edge detection methods that can be distinguished into three different categories:

The first category contains search-based methods that approximate derivatives of the image function using differences. The search-based methods detect edges by first computing a measure of edge strength, usually a first-order derivative expression such as the gradient magnitude, and then searching for local directional maxima of the gradient magnitude using a computed estimate of the local orientation of the edge, usually the gradient direction. These edge detectors (Kirsch, Pratt, Prewitt, Sobel) are based on convolution in very small neighborhoods and work well only for certain images. The main disadvantage of these edge detectors is their dependence on the size of the object and sensitivity to noise.

The second category contains zero-crossing methods that are based on the zero-crossings of the image function second derivative. The first derivative of the image function should have an extremum at the position corresponding to the edge in the image, and the second derivative should be zero at the same location. However, it is easier and more precise to find zero-crossing positions than an extremum.

The third category contains operators that attempt to match an image function to a parametric model of edges that will not be discussed in more detail. For our data set we used the Canny Edge detector that is based on the zero-crossing principle.

Canny Edge Detection

The Canny Edge Detection was developed by John F. Canny in 1986, and is a multi-stage algorithm to detect edges in images that are corrupted with white noise [Canny, 1986]. It aims to detect edges in an optimal way that means:

- Good detection: No important edges should be missed and the responses should be robust.
- Good localization: Distance between the actual and the located location of the edges should be minimal.
- One Response: Edges should be marked just once; This is partly covered already by the first criterion, since when there are two responses to a single edge, one of them should be considered as false. This criterion should also minimize image noise by eliminating false edges.

The algorithm consists of the following stages described in more detail below. The algorithm starts by convoluting an image f with a Gaussian of scale σ. It uses a Gaussian filter based on the first derivative, because it is susceptible to noise present in raw unprocessed image data. Assuming two-dimensional convolution, the image data are differentiated with respect to the directions x and y. The gradient of the surface of the convoluted image function in any direction is possible to compute from the known gradient in any two directions. From the computed x and y gradient values, the magnitude and angle of the slope can be calculated from the hypotenuse and arctangent.

$$G = \sqrt{G_X^2 + G_Y^2}$$

$$\theta = \arctan\left(\frac{G_Y}{G_X}\right)$$

With the rate of intensity change found at each point in the image, edges must now be placed at the points of maximum, or rather non-maximum points must be suppressed. A local maximum occurs at a peak in the gradient function, or alternatively where the derivative of the gradient function is set to zero. However, in this case it is preferable to suppress non-maximum perpendicular to the edge direction, rather than parallel to the edge direction, since the edge

strength is expected to continue along an extended contour. The algorithm starts off by reducing the angle of gradient to one of 4 sectors shown in Figure 35 "Gradient Sectors".

At each point, the center element of neighborhood is compared with its two neighbors along line of the gradient given by the sector value. If the central value is non-maximum, i.e., not greater than the neighbors, it is suppressed. From this stage referred to as non-maximum suppression, a set of edge points, in the form of a binary image, is obtained. These are sometimes referred to as "thin edges".

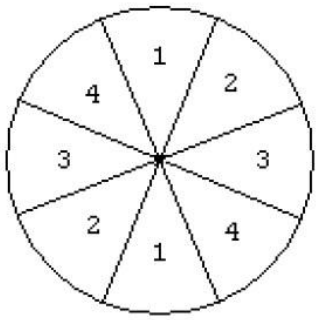


Figure 35: Different Gradient Sectors for edge detection

Large intensity gradients are more likely to correspond to edges than small intensity gradients. It is in most cases impossible to specify a threshold at which a given intensity gradient switches from corresponding to an edge into not doing so. Most thresholders use a single threshold limit, which means that if the edge values fluctuate above and below this value, the line appears broken. This phenomenon is commonly referred to as "streaking". The Canny operator uses the so called hysteresis thresholding. Hysteresis counters streaking by setting an upper and lower edge value limit. Considering a line segment, if a value lies above the upper threshold limit it is immediately accepted. If the value lies below the low threshold it is immediately rejected. Points which lie between the two limits are accepted if they are connected to pixels which exhibit strong response. The likelihood of streaking is reduced drastically since the line segment points must fluctuate above the upper limit and below the lower limit for streaking to occur. The final result of this processing step is a binary edge image of the true orthophoto where edges correspond to intensity changes within the RGB image.

Line Rasterization

In a second step we have rasterized the vector data of the cadastral information so that it matches with the data coming from the true orthophoto. The aim is to determine which points in an n-dimensional raster should be plotted in order to form a close approximation to a straight line between two points. To solve this problem we use a method introduced by Jack E. Bresenham in 1962, the "Bresenham's line algorithm" [Bresenham et al., 1965].

Feature Matching

Both are now binary raster images that can be matched. Matching is a key problem in digital image analysis, and edges are perhaps the most important low-level image features. If we compare binary images, by counting the number of identical pixels in the search image and the template, the total difference will only be small when most pixels are in exact agreement. Since

there is no continuous transition between pixel values when dealing with binary images, the distribution produced by a simple distance function will generally ill-behave. The problem with binary images is that even the smallest deviations between image patterns, caused by a shift, rotation or distortion, can create very high distance values. A simple distance function therefore wouldn't give any indication of how far away and in which direction to search for a better position. What we want to do is to find the optimal position where the template coincides with the search image. Therefore, we need a distance function that is smooth and more tolerant against small deviations between binary image patterns.

Distance function

Given a binary image I(r,s) = I(p) in a first step we calculate for every pixel the distance to the nearest foreground pixel (edges). We denote

$$FG(I) = \{p|I(p) = 1\}$$

$$BG(I) = \{p|I(p) = 0\}$$

as the set of coordinates of the foreground and background pixels. The distance transform of I, D(p) ϵ R is defined as

$$D(p) = \min_{p' \in FG(I)} dist(p, p')$$

In case that a pixel in the image I(p) is a foreground pixel itself, then the distance D(p) = 0, because there is no shift necessary to move this pixel onto a foreground pixel. The distance function dist(p,p') measures the geometric distance between two coordinates p = (r,s) and p' = (r',s'). This distance can be calculated using various functions like the

$$dist(p, p') = \sqrt{(r - r')^2 * (s - s')^2} \quad \text{Euclidean Distance}$$

$$dist(p, p') = max(|r - r'| * |s - s'|) \quad \text{Chessboard Distance}$$

$$dist(p, p') = |r - r'| + |s - s'| \quad \text{City Block Distance}$$

However, the calculation of the distance transform for every pixel in an image I is computationally expensive and difficult to implement; therefore, we use the so-called Chamfer matching that is an efficient method for the computation of the distance transform.

Chamfer Matching

Chamfer Matching, implemented after [Borgefors et al., 1988] traverses the image twice by propagating the computed values across the image in the form of a wave. It starts at the upper left corner of the image and propagates the distance values downward in a diagonal direction and the second traverses in the opposite direction from the bottom to the top. For each traversal, a distance mask is used for the propagation of the distance values.

$$M^L = \begin{bmatrix} m_2^L & m_3^L & m_4^L \\ m_1^L & x & . \\ . & . & . \end{bmatrix} \text{ and } M^R = \begin{bmatrix} . & . & . \\ . & x & m_1^R \\ m_4^R & m_3^R & m_2^R \end{bmatrix}$$

Both M^L and M^R describe the geometric distance between the current pixel marked with x and the relevant neighboring pixels depending on the used distance function. Chamfer Matching uses the distance transform to localize the points of maximum agreement between a binary target image I and a binary reference image R. It uses the accumulated values of the distance transform as the match score Q where at each position (r,s) of the reference image R, the distance values are accumulated corresponding to all foreground pixels in R.

$$Q(r,s) = \frac{1}{K} * \sum_{(i,j)\in FG(R)} D(r+i, s+j)$$

where K is the number of foreground pixels in R. If at some position each foreground pixel in R coincides with a foreground pixel in I, then the sum of the distance values is zero that indicates a match. The more foreground pixels of R that fall onto distance values greater than zero, the larger the score value Q is. The best match is found at the global minimum of Q.

$$P_{opt} = (r_{opt}, s_{opt}) = argmin\, Q(r,s)$$

The complete algorithm to compute the matching score Q is summarized in Figure 36. The result of this processing step is a more accurate overlay of the true orthophoto and the cadastral information by reducing discrepancies in the two input images.

```
ChamferMatch (I, R)
        I: binary image of size w_I × h_I
        R: binary template of size w_R × h_R

    Step 1 – initialize:
    D ← DistanceTransform(I)                          ▷ Alg. 17.1
    K ← number of foreground pixels in R
    Set up a match map Q of size (w_I − w_R + 1) × (h_I − h_R + 1)

    Step 2 – compute match function:
    for r ← 0 ... (w_I − w_R) do              ▷ set origin of template to (r, s)
        for s ← 0 ... (h_I − h_R) do
            Evaluate match for template positioned at (r, s):
            q ← 0
            for i ← 0 ... (w_R − 1) do
                for j ← 0 ... (h_R − 1) do
                    if R(i, j) = 1 then       ▷ foreground pixel in template
                        q ← q + D(r + i, s + j)
            Q(r, s) ← q / K

    return Q
```

Figure 36: Processing Steps of Chamfer Matching Algorithm [Burger et. al 2008]

5.4. Minimizing Building Overhangs and Mismatches

The image segmentation result is in the same coordinate system as the orthophoto. Therefore, the cadastral map could be used directly to cut a classification map into data per property. However, a look at the result after Chamfer Matching reveals that the cadastral boundaries and the building classification don't match perfectly.

Classification uncertainties are caused by mismatches between cadastral line work and the building outlines when a cadastral boundary line coincides with a building facade. We may find façade details such as balconies, or roof extensions in the form of eaves that violate the cadastral delineations. The main reasons why we have these mismatches are building and roof overhangs that occur in a 2.5D DSM that affect the building classification. Other reasons are misclassifications of the segmentation image. A solution would be the use of a fully 3 dimensional DSM that already considers and removes these overhangs. Having the cadaster available offers the option of changing the classification along building and property boundaries.

While we assume that these small differences will be insignificant in the application to a property characterization, it is important that we understand and hopefully reduce or remove them.

A roof's eave causes a small parallel shift vis-à-vis the cadastral data to make them fit the segmentation result. The inconsistencies from bay windows or balconies would also disappear if one were to redraw the classification boundaries. These interventions would affect the classification results and rate of misclassifications. These considerations create the need for methods to automatically improve the alignment of the cadastral line work and the segmentation boundaries. We want to address three different possibilities to minimize these misclassifications:

The first solution is to perform this improvement of the building classification by hand. Therefore, the overhanging building parts are eliminated by overlaying the cadastral boundaries and the building classification.

Another possible solution works semi-automatic. Linear areas that typically occur at the borders of the building classification due to the overhangs are detected and removed. Therefore, all generated building parts are evaluated depending on the following parameters: the minimum area and the extent of the rectangular boulding box that envelops a certain building part. With these parameters most of the building overhangs can be eliminated. However, a visual inspection of the result is always necessary to check the detected removable areas. Moreover, this method completely ignores the visual information coming from the aerial images. It is just based on the result of the overlay of the cadastral data and the building classification. If there are errors within the segmentation or in the cadastral data these errors are neglected.

The third alternative works fully automatic and uses a plane sweep to reconstruct the building facades in 3D. It also projects the reconstructed 3D façades on the xy-plane. This method has a substantial advantage compared to other methods. The result of this computation delivers the exact building footprint that can also be used to generate a fully 3D DSM by including all

potential information from the aerial images. A more detailed description of this method is given in chapter 7.

5.5. Building Façade Separation

However, sometimes building boundaries derived from the cadaster may be insufficiently accurate, for example due to old databases with lower accuracies or inaccuracies due to transformation between two coordinate systems. For this reason, it can happen that a cadastral boundary coming from an old map is displaced by up to several meters [Feucht, 2008] and therefore divides two buildings incorrectly. Attached buildings that are located on one single property present another problem (see Figure 37).

Figure 37: Overlaying the cadastral map (highlighted in red) over the true orthophoto shows displacements of the cadastral boundaries or missing cadastral boundaries vis-à-vis the photography. In this case the cadastral boundary between these two buildings is missing.

To overcome this problem we incorporate the information from vertical aerial images to separate building façades [Hammoudi et al., 2011; Wendel et al., 2010]. We employ a method introduced by [Wendel et al., 2010] that is able to separate building façades using multi view stereo. Separated façades can then get analyzed for their details. In this previous project, the source material consisted of a set of overlapping [thus redundant] images using a moving vehicle and calibrated automated cameras.

Within this project, we adapt their method to vertical aerial images in the hope to increase the accuracy of a building block separation beyond that obtainable from previous approaches. We

determine the building block outlines using the building classification and use the height values from the DSM to determine the approximated "façade quadrilaterals". We also incorporate height discontinuities using the height profiles along the building outlines to enhance our façade separation. In a next step, we detect repeated patterns in these "façade images" and use them to separate the building blocks of the façades from one another. In order to produce accurate results one needs (a) a DSM with well-defined building roof lines to avoid ragged building edges (see Figure 20), as well as (b) a precise Classification image from the test area (see

Figure 21).

Proposed Method

The proposed method consists of three steps. First, we extract the building block outlines using the classification image. To generate the necessary straight lines we apply a recursive line simplification scheme on the building block footprints. In a second step, the height values coming from the DSM get assigned to the extracted building block outlines. In the original façade separation algorithm this additional information is not necessary; however, for vertical aerial images it is crucial for the outcome. After this is done, all façade strips are projected into the vertical aerial images and rectified. In a last step, the façade separation is performed on the rectified façade strips.

Building Block Outlines

For each building block we have to determine its outline. The building objects obtained from the image classification are an approximation of the intersection of a façade with the ground. The goal is to isolate the contour of each building block. Initially, this contour is in the form of pixels in need of a vectorization. This has been studied for a long time and a choice of different methods exists. In our case, we employ the recursive line simplification by [Douglas et al., 1973]. Figure 38 illustrates the result of this calculation for a subset of the Graz test area. In this area, 23 straight lines were extracted. A more detailed discussion about line simplification algorithm is conducted in chapter 6.

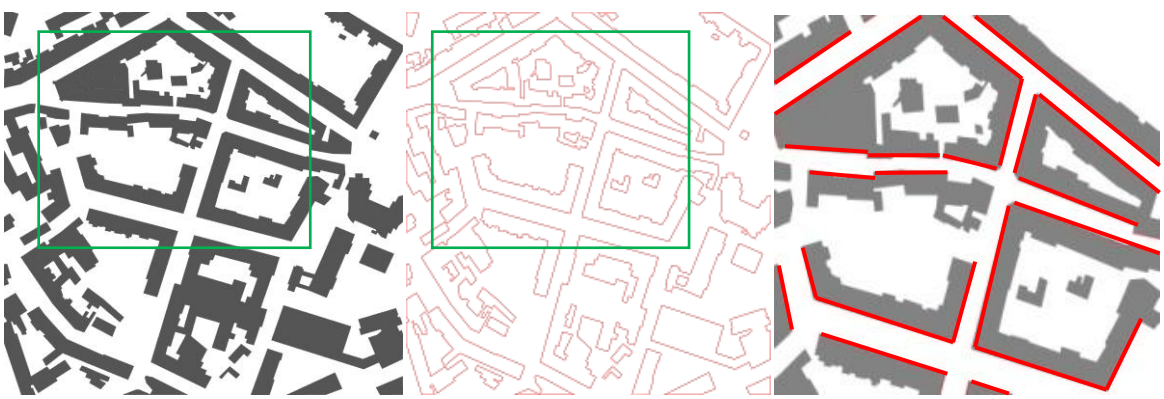


Figure 38: The classification layer "*building*" is based on color, texture and elevation values. (a) is the binary layer; (b) presents the contours as a raster image and (c) are the detected façade footprints for the area within the green rectangle highlighted in red

Incorporating Height Data

Because of the low resolution of the building façades, this step is crucial for the outcome of this approach. The incorporation of the height values coming from the DSM has many benefits. First, we need it to add the third dimension to the 2D building block contours to determine the façade strip quadrilaterals. In the simplest case, we assign one height value to the previously defined 2D lines. The result is a simple façade quadrilateral that can be projected into the aerial images (see Figure 39).

Second, using the elevation values with the 2D building contour allows us to split buildings with different heights in advance. One may determine a measure of the building block symmetry for the elevations along the footprint: if façades get associated with different building heights, one may have reason to break the previously defined line into its parts. In our case, these lines are split by calculating the gradient of the respective height profile (see Figure 40). If the gradient exceeds a certain value, buildings are getting separated. These locations are stored and used as additional information for the following façade separation approach.

(a)

(b)

Figure 39: ´vertical aerial image showing façade strip and superimposed detected façade polygon marked in green (a) Projected façade quadrilaterals without the use of height information and (b) projected façade polygon by incorporating height information.

And third, the use of the height data allows us to segment façades in the vertical aerial images in more detail. Instead of four corner points for each façade strip, we get a polygon that determines the outlines of the single building façades. This is helpful to enhance the separation results, especially when the heights of attached buildings vary substantially (see Figure 40).

After the generation of the façade strip quadrilaterals and the additional information coming from the elevation data, we can project them into the aerial images and rectify them. Figure 39 illustrates a projected façade strip without additional height information and the enhanced version by incorporating different buildings heights.

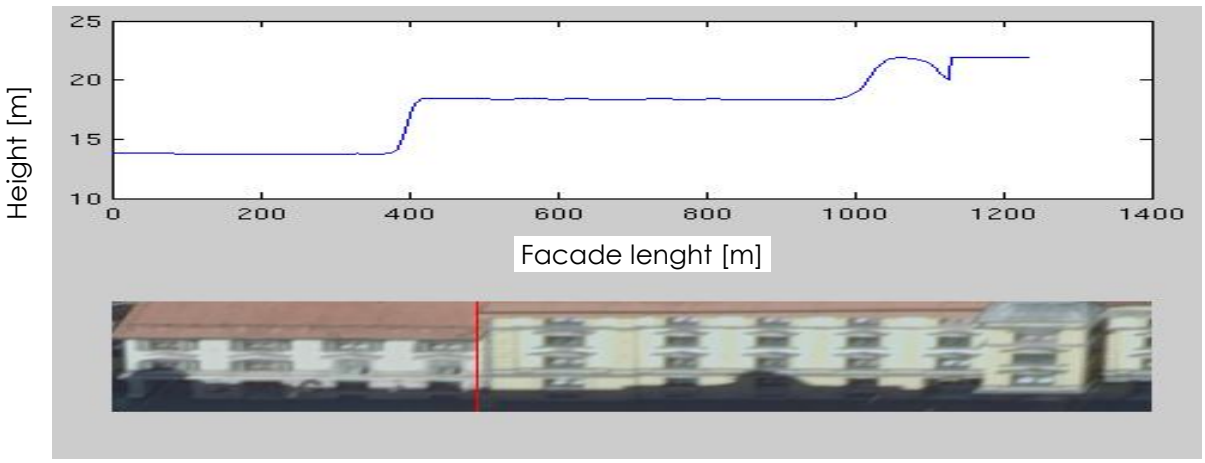


Figure 40: Above -- Height profile of a façade strip; Below -- determined façade splits (highlighted in red) based on height discontinuities visible in height profile

Façade Separation and Segmentation Algorithm

In this section, we describe the façade separation algorithm in more detail. The result of the previous step produces façade areas, not individual façades per building. To be able to characterize the single buildings we have to split these façade areas into single façades. The applied algorithm was introduced by [Wendel et al., 2010] and consists of two major steps: First, repetitive patterns are detected in the façade images. In a second step, the resulting pairs of interest points are then used to separate the results.

Finding Repetitive Patterns

In a first step, repetitive patterns in the images get associated with façades. The method uses Harris corners as interest points. Figure 41 illustrates the detected Harris corners for one façade area [Harris et al., 1988] using street-level imagery for illustration. In a next step, the color profile on a straight line between every interest point and the 30 nearest neighbors is calculated. The color profiles are constructed using a 20-dimensional normalized descriptor for each of the three colors RGB, thus totaling 60-dimensions. Finally, the 60-dimensional descriptor is normalized. A kd-tree method is then used for matching the descriptors. Matches with more than 10 descriptors are eliminated in advance because they are not discriminative enough. In a last step, the repetitive patterns are located in a voting matrix. Within this voting step all matching profiles vote for the similarity of the respective pair of start- and endpoints. The results are stored in a list of contributing profiles for every possible pair. To strengthen the results of repetitive patterns on a

façade, we have eliminated all matches between two profiles that are too far apart or too close to each other.

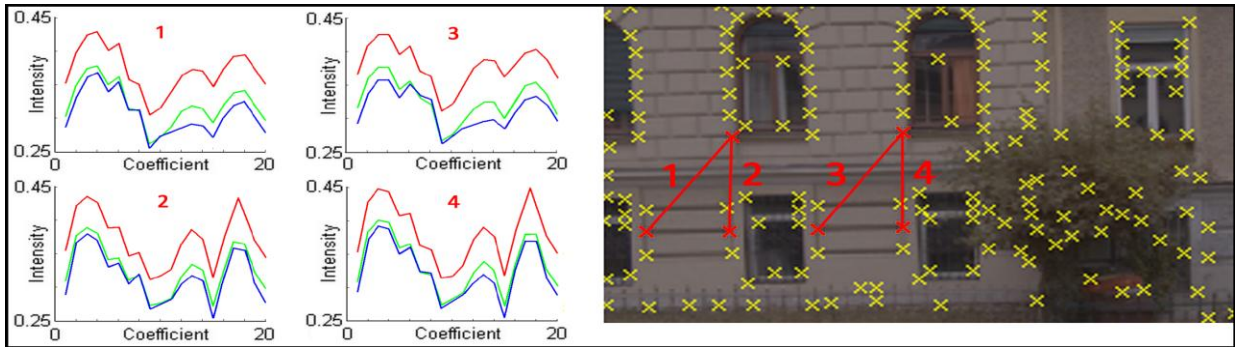


Figure 41: Detected Harris corners and the extracted intensity profiles consisting of 20 values for every RGB channel; taken from [WENDEL ET AL., 2009]

Façade Separation

In a next step, the processing of the single façade is discussed in more detail. Due to the natural settings of objects in these images, we assume that repetitive patterns occur along the horizontal direction (Figure 42b) and the separation of the façades occurs in the vertical direction. Therefore, the lines between the matched interest points are projected into the horizontal axis constructing a match cost histogram (Figure 42c). The match count is normalized to obtain the percentage of all matches, the repetition likelihood (Figure 42d). Then the façades are segmented by determining a separation area (area where one façade ends and the next begins). This is done by defining areas with a low likelihood as separation areas and areas with a high likelihood as repetitive areas. To be able to determine the exact split between two façades in a last step we look for the global maximum in these areas. The results of the different processing steps are demonstrated in Figure 42.

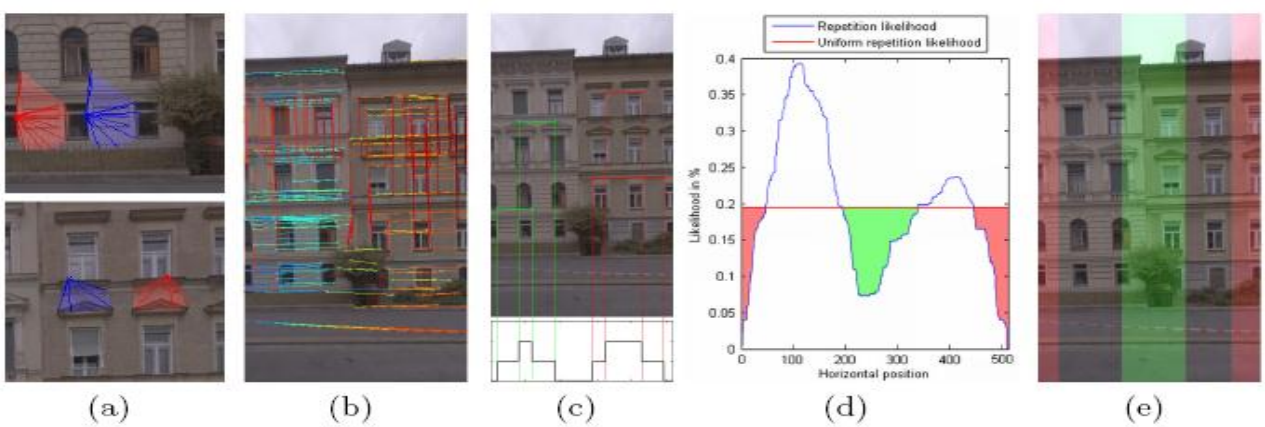


Figure 42: From StreetSide data to separation: (a) Matching of arbitrary areas (b) Detected repetitive patterns (color coded lines) (c) Projection results in a match count along the horizontal axis (d) Thresholding the repetition likelihood with the uniform repetition likelihood (e) Resulting repetitive areas, separation areas (green), and unknown areas (red);taken from [WENDEL ET AL., 2009].

The results are the locations for possible splits of building facades. In a final step, we calculate the normal to the previously determined façade footprint by incorporating the generated façade splits. Along these normals we finally can split the single buildings.

5.6. Data per Property

Once the orthophoto and cadastral parcel match or the façade separation is completed, one can proceed to cut all data sets along parcel boundaries. All data is in the same coordinate system; therefore, the cadastral map and the generated building separations can be used directly to cut the building classification, the DSM and the DTM into data per property. Cutting these data sets along property boundaries is trivial and based on the cadastral data after Chamfer refinement and the removal of all overhangs. A per-property presentation of the various map data is shown in Figure 43.

Elevation values associated with the classification layer "building" will provide a visual impression of the building shape. Similarly, elevations associated with a vegetation layer will describe the shape of trees. However, in those data sets, the property boundaries do not play any role. This partition into different land use categories (see Figure 43) is of special interest for urban planning and affects the value of properties.

(a) Orthophoto (b) DSM (c) Vertical Objects (d) Classification

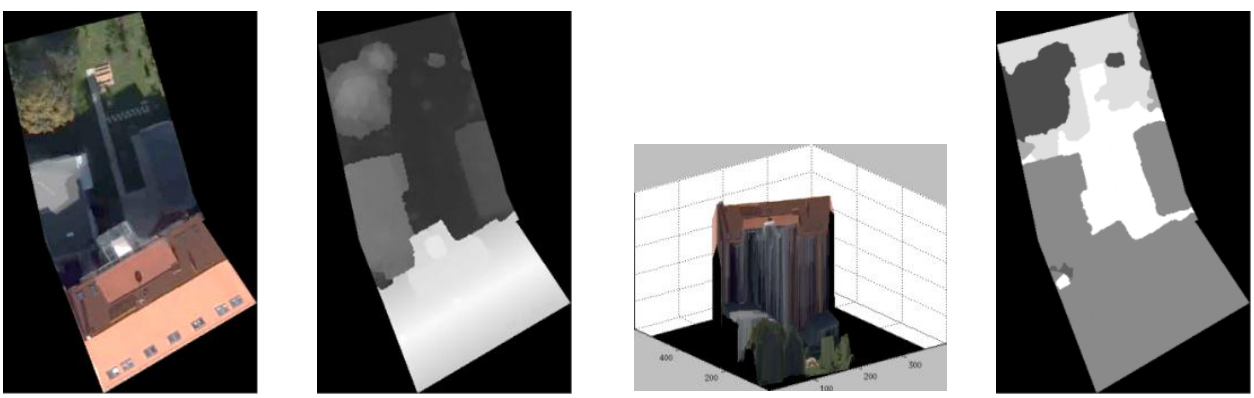


Figure 43: sample property and the source data for that property. (a) the orthophoto segment (b) DSM elevation data (c) the elevations on top of the Bald Earth. (d) shows the classification result with seven object classes.

5.7. Experiments

All the experiments in this section were performed for the Graz test data set. This section is divided into three parts. The first part covers the evaluation of the used source data, in our case the classification image and the cadastre. The second part evaluates the accuracy of the coordinate transformation (the shift between the cadastre and the true orthophoto) and the result after Chamfer Matching. Therefore, we used different input data like the DSM, the true orthophoto and the building classification and compared the results. We also analyze the

proposed methods for the removal of roof and building overhangs using both the semi-automatic method and the fully automatic method. The last part of the experiments deals with the evaluation concerning the proposed façade separation method that is necessary when cadastral boundaries are not useful to extract the single buildings.

Accuracy of Segmentation and the Cadastre

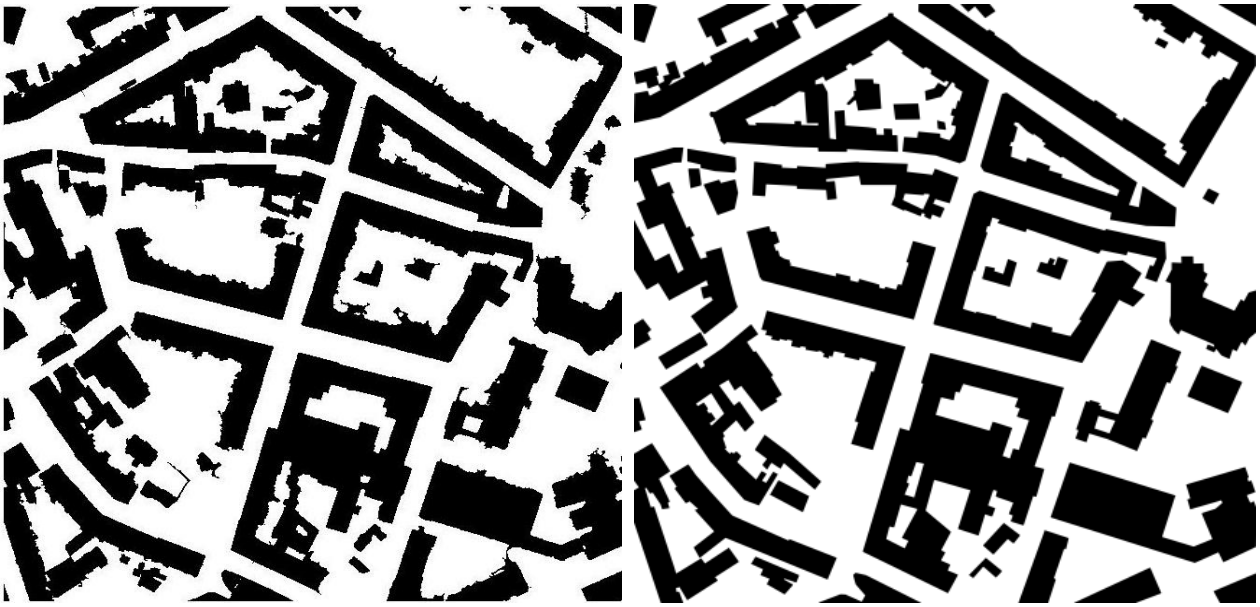
An evaluation of the automated segmentation is based on manually developed ground truth. Therefore, we used the True orthophoto and digitized the building contours including roof overhangs. Figure 44ab shows the binary mask of the determined ground truth.

For the evaluation of the segmentation, we overlaid the automatically segmented buildings with the ground truth data and calculated the difference image. Figure44c illustrates the result of this overlay. Red areas represent correctly determined buildings, blue areas are false positives and green areas are false negatives caused e.g. by cars and vegetation. Table 19 shows the result of this analysis.

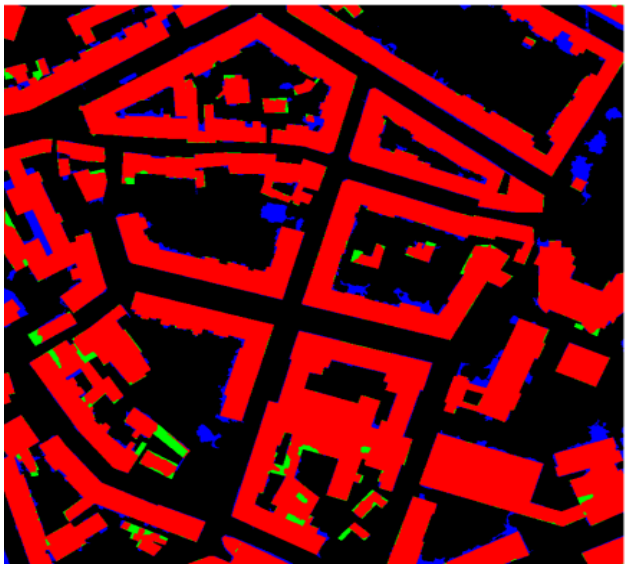
	Total	*False positives*	*False negatives*	*Difference*	*Detection rate*
Buildings [pixel]	6.726.617	556.802	185.027	5.984.788	89%

Table 19: Evaluation of building classification based on pixels

For the example of the building information, this experiment confirms that the classification approach implemented by [Gruber-Geymayer et al., 2006b] and [Zebedin et al., 2006] delivers an accuracy of 90%. In over 90% of the cases, classification uncertainties are caused by vegetation. Almost all of the false negatives (95%) occur because of trees covering building roofs. False positives are caused because of trees (75%) and other variables (like cars and smaller objects) being located next to a building.



(a) building classification (b) manually determined ground truth



(c) difference image

Figure 44: Evaluation of building classification (a) computed building classification (b) manually determined ground truth (c) difference image (red: correctly detected buildings, blue: false positives, green: false negatives)

Discussion

We could demonstrate that vertical aerial photography holds considerable promise as a source of descriptive information about individual properties. Such information is expected to be useful in Internet-based mapping systems. Measures of surface areas covered by a parcel, its buildings and other land use categories are obvious results from vertical aerial mapping, and are available at an accuracy of about 90% of the surface area. These 10% errors result from uncertainties of a parcel's definition in an aerial image, given a cadastral record. Uncertainties also exist in the automated classification of the contents of the image and the areas covered by buildings, vegetation, and various land use categories. The question now is: "Are these results good or not?" The answer to this question depends strongly on the application. For instance, for applications that deal with land use these results are not bad. In that case, the precision of the building contours is not very relevant. For our purposes, the precision of the building contour is crucial for the delineation of the façade footprints. We will provide an answer to this question in the course of this project.

Matching the Cadastre with the True Orthophoto

In the next part, we want to analyze the results from the coordinate transformation, the following Chamfer Matching and the removal of building and roof overhangs.

A 2-step process serves to match the 2D cadastral vector map (in its own coordinate system) and the 2D orthophoto (in its different coordinate reference). In a first step, the cadastral point coordinates simply get converted from their *Gauss-Krüger M34* values to the building

classification's *Universal Transverse Mercator UTM* system. In an ideal world, this would solve the registration problem. It does not, as illustrated in Figure 45 (left).

Figure 45: Overlaying the cadastral map over the orthophoto will leave some small errors that need to be found and removed. Left is the overlay after step 1 of the registration, namely the coordinate transformation from the Gauss-Krüger system to the UTM system. To the right is the result after a Chamfer- match has been applied.

We have performed the coordinate transformation independently for four regions in the Graz test data set (Jakomini, Kronesgasse, Elisabethinergasse, Puntigam) and achieved in all four regions constant offsets. Table 20 summarizes the results for the four different test areas. The results of table 20 shows a quite constant offset in x-direction and a small offset in y-direction and no rotation which would have been caused by an error within the coordinate transformation.

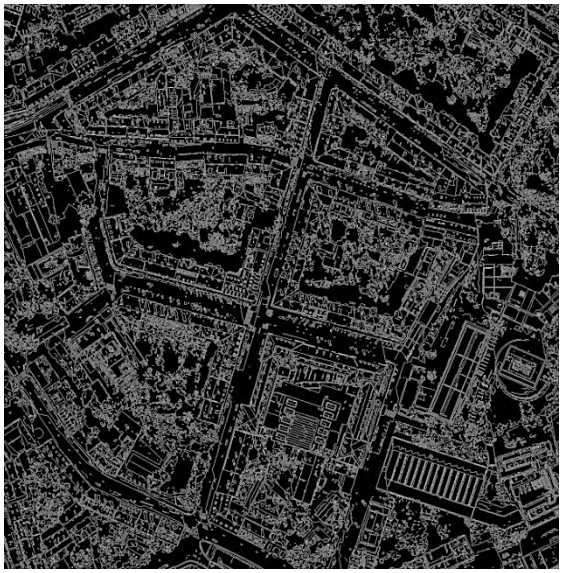
Area	*x-Offset*	*y-Offset*
Kronesgasse	+ 7	+ 1
Jakominiplatz	+ 6	+ 2
Elisabethinergasse	+ 7	+ 2
Puntigam	+ 5	+ 3

Table 20: Evaluation of Coordinate Transformation; Mismatches between cadastre and orthophoto in pixels (GSD 10cm)

Such mismatches can be the result of the different histories of the cadastral data and their focus on 2D local information or inaccuracies coming from GPS and IMU measurements. We do not allow for a local deformation of the cadastral data. Instead, the cadastral maps are treated as rigid 2D entities where changes are only permitted in translation.

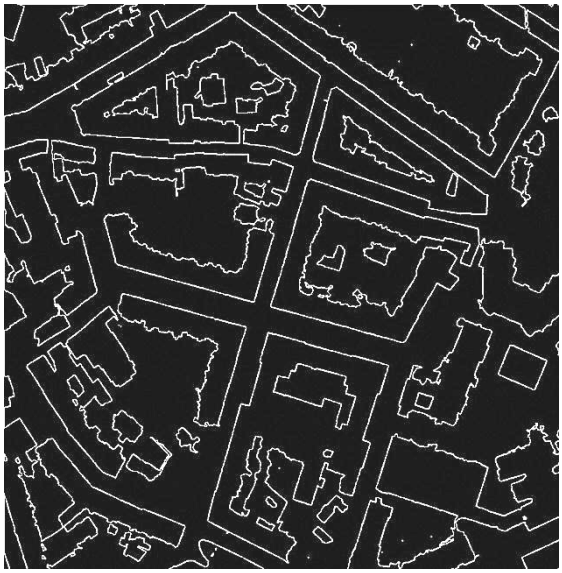
Chamfer Matching

We apply chamfer matching to conflate the vector-type parcel data with the rastertype Orthophoto. This is applicable if sufficient image information is available to define the parcel boundaries by natural features. Major parcel-vector matches with imagery are along street outlines and where fences exist. We also analyzed the use of other imaging products instead of the true orthophoto for chamfer matching like the DSM and the building classification.



(a) Orthophoto

(b) DSM



(c) Classification Image

(d) Distance Image

Figure 46: Input data to the chamfer match. Edge detection results for (upper left) True Orthophoto (upper right) DSM (lower left) Building Classification and (lower right) calculated distance image from cadastral data.

Figure 46 shows the detected edges for these data. The Chamfer Matching algorithm converts the vector image of the cadastral map into a raster image and assigns to each pixel a grey value as a function of the distance from the nearest vector. This is denoted as a distance image. The resulting raster image from the cadastre is shown in Figure 46.

We have performed the chamfer matching based on all four input data sets (true orthophoto, DSM and building classification) for all four test areas in Graz and achieved for all three data sets similar results (see Table 21). The best results were achieved using the building classification where we could reduce the mismatch from ±7 to ±1 pixel. For the DSM we achieved an improvement from ±7 to ±2 pixels and for the true orthophoto from ±7 to ±3 pixels. The reason why the true orthophoto produced the least accurate results is due to the heterogeneity of the image data. The results show that we could successfully remove all shifts. For the further evaluation, we use the result obtained from the building segmentation.

	Original data		***True Orthophoto***		***DSM***		***Building seg.***	
Area	*x-offset*	*y-offset*	*x-offset*	*y-offset*	*x-offset*	*y-offset*	*x-offset*	*y-offset*
Kronesgasse	+ 7	+ 1	+ 3	+ 1	+ 2	+ 1	+ 1	+ 1
Jakominiplatz	+ 6	+ 2	+ 2	+ 1	+ 2	+ 1	+ 0	+ 1
Elisabethinergasse	+ 7	+ 2	+ 3	+ 1	+ 1	+ 0	+ 1	+ 1
Puntigam	+ 5	+ 3	+ 2	+ 2	+ 2	+ 1	+ 1	+ 0

Table 21: Evaluation of Chamfer Matching: Offsets before and after Chamfer Matching using different input data (in pixel)

Removal of Roof and Building Overhangs

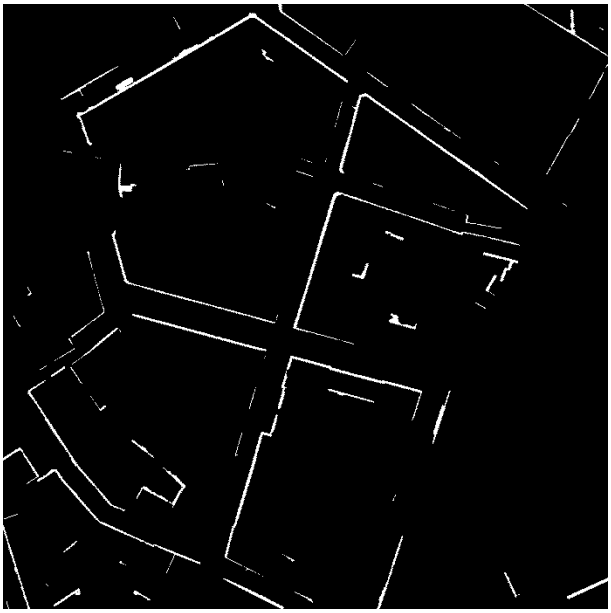
In the next part, we want to explain the experiments we did for the removal of the roof overhangs and building overhangs. In our test area, façade details such as balconies, or roof overhangs in the form of eaves, violate the cadastral delineations. We have performed an evaluation for two proposed methods as explained in the first part of this chapter:

- Semi Automatic Removal (elimination of linear and small areas)
- Fully Automatic Removal (plane Sweep)

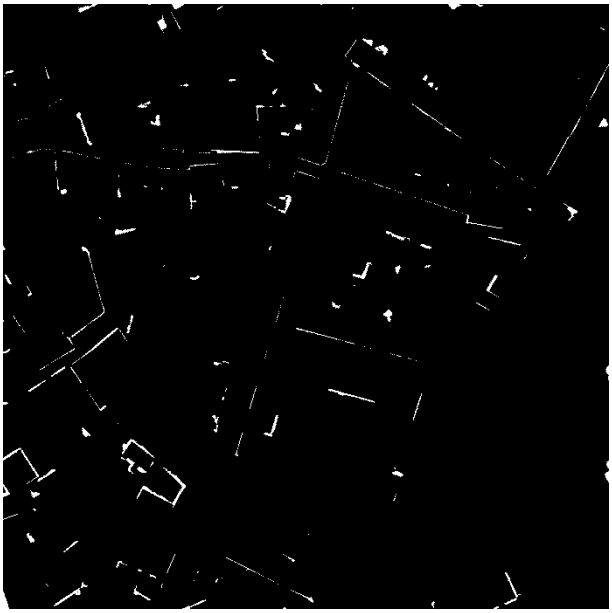
In the first experiment, we have defined two parameters to analyze the result after splitting the building classification using the cadastral information. These parameters are defined as the minimum size that a resulting object can have, and the extent that defines the proportion of the pixels in the bounding box that are also in the region.

These detected overhangs are now incorporated in the building classification and are removed. The original building classification with overhangs, misclassifications and the resulting are shown in Figure 48. In all four test areas in Graz, no manual removal of overhangs was necessary.

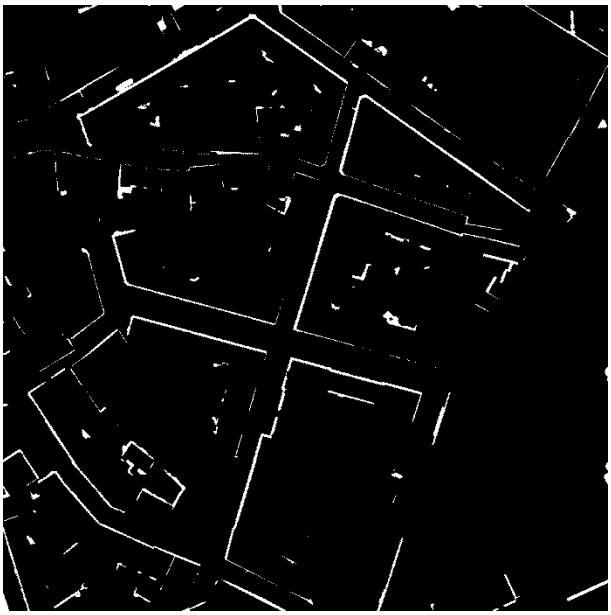
Figure 47 illustrates the detected areas for each parameter and the combination of these results. These detected overhangs are now incorporated in the building classification and are removed. The original building classification with overhangs, misclassifications and the resulting are shown in Figure 48. In all four test areas in Graz, no manual removal of overhangs was necessary.



(a) Mimimum Area Criterium



(b) Extend Criterium



(c) Combination of a) and b)

Figure 47: Detected Building and Roof overhangs (a) applying minimum area parameter (b) applying extent parameter and (c) combination of both results.

Figure 48: Overlay of building classification and cadastre. Left is with the discrepancies due to roof eaves and façade detail, to the right is the cleaned-up version.

The second method to eliminate roof and building overhangs is based on plane sweeps. Thus facades are reconstructed in 3D and the resulting point cloud is projected onto the xy-plane. A detailed evaluation of this proposed method is given in chapter 7.

Discussion

The cadastre-based separation of the images, classification results and elevation data is fundamental to achieve the description per property. In the absence of a well-developed and accurate cadastral system, one would have to devise an approach with street maps and address points. Each address point would be a seed pixel in the imagery from where to grow the region of a property. The street map would delimit the properties vis-à-vis public spaces. Properties would have to be defined from these approximations and rely on image content. Cadastral information without a cadastral survey would be desired and represents a future research challenge.

Evaluation of Façade Separation Method

The 216 buildings from the Graz test area are grouped into 65 major building blocks with different sizes and shapes (see Figure 38). Because some of these blocks consist of just one single building in a pre-processing step all these blocks were eliminated. Furthermore, just façade areas with a minimal length were used for this evaluation. We performed the evaluation of the proposed method for 31 façade images containing a total of 121 single façades using three different settings.

Façade separation without elevation data

For our first experiment, we used only the vertical aerial images without additional height information to determine the impact of this information on the outcome of the processing. In this

case, we used the RGB-information coming from the aerial imagery and one height value for the whole façade area (see Figure 49). We performed the calculation for 31 façade images and achieved a success rate of 65% that corresponds to 79 detected building façades out of 121. Figure 49 shows the result using these settings. Reasons for the low detection rate are the low resolution of the aerial images that handicaps the extraction of repetitive patterns.

Façade separation including elevation data

In our second experiment, we incorporated height information in our façade separation. This time preliminary facade splits are also included. Figure 40 illustrated an example of such a detected façade split. Like before, we performed our experiments for 31 façade images and achieved an increase of the success rate from 65% to 88%; this corresponds to 106 detected building facades. Figure 40 and Figure 49 show the results for façade separation, including this additional information. As one can notice, all façades could be separated correctly.

In our experiments, we had 21 false positives that occurred due to the nature of the used method (see Figure 49). This is due to irregular facade texture that leads to an undesired separation of one facade.

(a)

(b)

(c)

Figure 49: Result of façade separation (green and red lines are detected splits) (a) without height information, (b) including height information and (c) façade with false positive marked in red

Façade separation using street side images

The question is: "*Are these results at 88% good or not?*" To be able to answer this question, we consider façade detection in high resolution and highly overlapping street side images. Tests with street side imagery have been performed in a subset of our test area with 9 separate building façades shown in 20-50 overlapping photos. The street-side images are taken in a forward look so that the façades are shown under an oblique angle. This helps in evaluating the influence of the perspective distortion. A detection rate was achieved of 97% [Wendel et al., 2010], whereby the facades were plane and best case.

Discussion

We need to compare the 88% success from vertical aerial images with the 97% success in street side images. The difference is far less than expected when one considers that the vertical aerial imagery nominally has 10 cm pixels, but a very oblique viewing perspective of façades, whereas the street side imagery is in the 1 to 2 cm range and looks far less obliquely at façades. Therefore, we conclude the façade separation from vertical aerial images to be feasible and successful.

CHAPTER SIX

CHARACTERIZATION OF REAL PROPERTIES

Contents

6.1	Introduction	135
6.2	Building Characterization	136
6.3	Experiments	142

6.1. Introduction

As a result of the data segmentation per property one obtains several important features: land area, areas used up by buildings, vegetation, water bodies and impervious surfaces. We have previously established that the areas are at an accuracy of within 90% largely due to limitations of the classification process. By introducing the 3rd dimension in the form of the dense point cloud, one will add relevant property information and improve the accuracy of area measurements previously obtained in the 2D classification.

We split the characterization into three major parts:

1. The characterization of buildings
2. The characterization of real properties without buildings (not discussed within this project)
3. The interaction of properties, e.g. regarding views and occlusions (not discussed within this project)

An overriding role is associated with the buildings; therefore, they are the focus of our efforts. The description of a real property consists of a table with coordinates and numbers. These define the number of buildings, type of building from a stored list of candidates, building height and footprint, number of floors, number and types of windows, presence of a basement floor, type of attic, roof type and roof details (such as an eave, skylights, or chimneys) the presence of a garage and its size as well as exposure to the sun and effects of shadows, the quality of views from a window etc. The characterization of building facades and building roofs are of great interest in our research. In greater detail, we will discuss facades in chapter 7 and roofs in

chapter 8. However, to be able to discuss facades and roofs in the next chapters, there are some processing steps that need to be explained.

Nevertheless, an entire characterization of the single properties also includes vegetation and land use characteristics. Interesting characteristics are the types and extent of impervious surfaces such as: a driveway and parking spaces, and descriptions about the type and size of the vegetation, the presence of a water body or the existence and type of a fence. However, these areas were beyond the scope of this project and therefore are not discussed.

6.2. Building Characterization

In chapter 5, we defined what a building is and how we can extract it from the input data. So far, we divided the building blocks into single buildings per property. In a next step, we want to characterize their details like facades and roofs. Therefore, we have to extract the outlines of a building that is crucial for the characterization of the single façades. We use a three-step approach to determine the location and size of a building facade. First is the need to find the façade's footprint, second is the assignment of a façade height and third is the projection into the respective aerial images.

Computing Footprints of Facades in 2D

We have two information sources for building footprints. One is from the image classification of roofs. The other is from the vertical elements of the DSM. The classification is typically based on color and texture, not however, on the 3D information of the DSM. Therefore, the two information sources are independent.

Figure 50: The classification layer "building" is based on color and texture. The left shows the binary layer, center is its contour in raster and finally at right is the geometric figure of the footprint (single lines are marked in different colors).

Building Contour using the classification layer building

The contours of building objects from the image classification approximate the intersection of the façades with the ground (see Figure 50). Initial contours need to get vectorized. Often the resolution of a contour is much higher than needed. For the sake of efficiency, we look for algorithms that can extract essential features from the detailed data of the building contour. These contours are getting simplified by minimizing the number of vertices, sufficient for a correct representation of the object. There are many different algorithms to simplify lines for different

applications [Douglas et al., 1973; Jenks, 1981; Lang, 1969; Reumann et al., 1974; Saalfeld, 1999; Tobler, 1964; Wessel, P. et al. 1996]. Such an algorithm consists in constructing a polyline with line segments larger than an accepted tolerance. This can be done by discarding recursively the subsequent vertices whose distance from a prior vertex is less than a certain maximum distance. A vertex that is farther away than that maximum distance is accepted as part of the new simplified polyline and becomes the new initial vertex for further simplification [Sunday]. For the detailed study of mathematical similarity and discrepancy measures, the Douglas-Peucker algorithm is pointed out as the most visually effective line simplification algorithm [Douglas et al., 1973; Ramer, 1972]. For this reason, we have used it for our purposes. It replaces the contour pixels P_i in a piecewise linear curve according to a specific tolerance d_{max}.

$$d_{max} = \max_{i=2\ldots n-1} d(P_i, \overline{P_1 P_n})$$

The contour pixels get replaced by straight lines, each defining one facade. The result consists of straight line segments as illustrated in Figure 51.

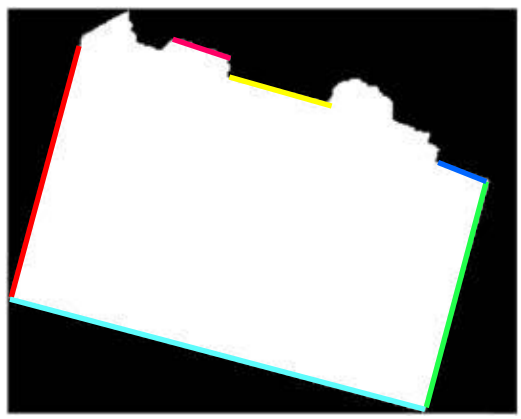


Figure 51: Determined line segments based on the building classification. Colored lines are detected line segments.

Building Contour using the Vertical Elements of the DSM

A second independent information source for a façade is the DSM. The first derivatives of the Z-values along X-rows and Y-columns of the dense point cloud represent the surface tangents. Where this is vertical, one will have a façade point. The proposed algorithm consists of the following steps:

- Loading the height-postings H_{ij} of the parcel for all rows i and columns j;
- Calculation of the first and second derivative H_{ij}' andH_{ij}'' of the height data H_{ij} in each line and column;
- Locating the maximum 1st and 2nd derivatives H_{max}'and H_{max}'' in each line and column, delivering candidate footprint postings;

- In a neighborhood around a candidate footprint location, determine the associated height H of a structure;
- Decide on valid footprint positions from the verticality of the DSM expressed by the values of H', the curvature expressed by H'' and building object expressed by the height H.

The positions of candidate footprint pixels are now in the raster format. We again convert this to line segments as in (a) above. The information can now be fed into the computation of a geometric figure of the building footprint as previously described. This geometric figure is the resulting "building mask" (see Figure 52). Other vertical objects may be trees, and those will also produce candidate footprint pixels. However, there exist straight line segments to replace those pixels causing those footprint pixels to be deleted.

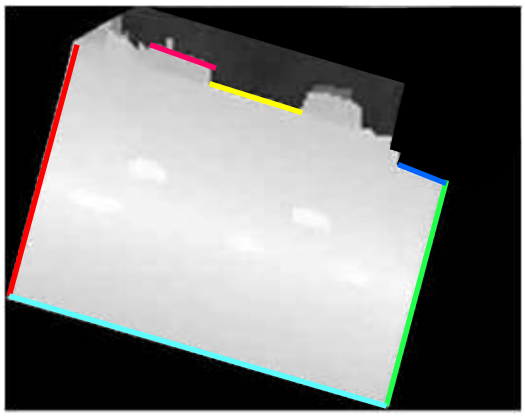


Figure 52: Determined line segments marked in different colors based on Vertical Elements in the DSM

Reconciling two Façade Data Sets: Segmentation Contours and DSM-Derivatives

The façade footprints from the image classification based on 2D color and texture need to be reconciled with the footprint from the 3D point cloud of the DSM. One initial approach is to simply use the mean between the two largely independent measures. Therefore, the two results are overlaid and combined. The main advantage of using both source data sets is that misclassifications can be reduced by incorporating elevation data coming from the DSM. The result of this merging of the two datasets are straight line segments shown in Figure 53.

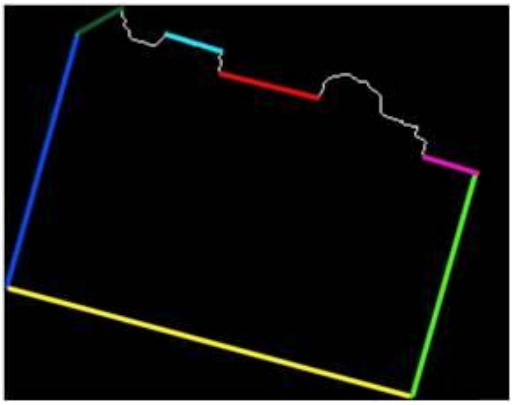


Figure 53: Merging of line segments from the two datasets shown in Figure 51 and Figure 52.

Façade Footprint Simplification

At issue is the number of façades a building has, which determines the building's complexity. The simplest building will have 4 facades at right angles. However, there is no limit to the complexity of building shapes and thus of the number of façades, and also of the angles formed between adjacent facades.

Without this simplification step it would yield to a quadruple number of facades, and each image would only show a small element. One will have to cope with ambiguities because of the smallness of the façade elements. The biggest problem is that the building classification that is used for the determination of building footprints is not accurate enough (see Figure 54) due to misclassifications.

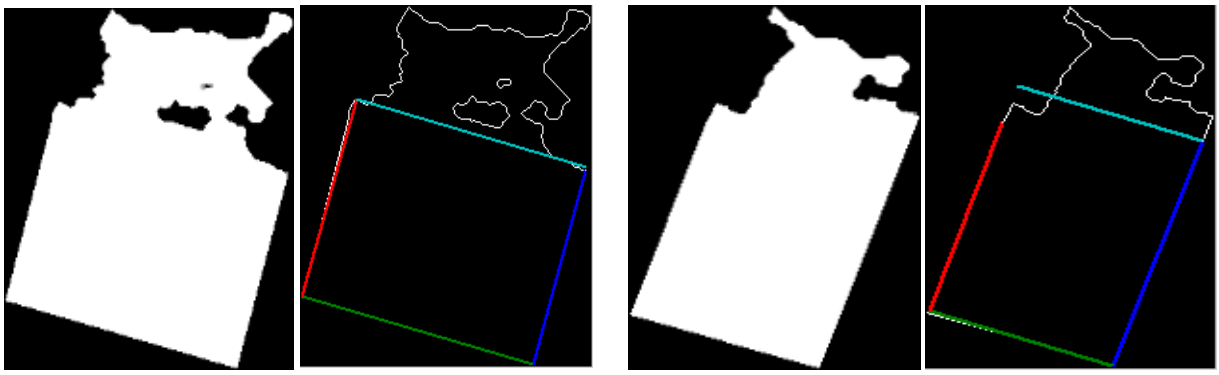


Figure 54: Building Footprints of two buildings from our test area. In these cases building classification is wrong and would lead to incorrect results when characterizing the facades.

We fit into the pattern of façade footprints a series of predefined shapes of building footprints. In the process, one hopes to develop a set of separate non-overlapping basic building objects or "sub-buildings". This is obtained via a segmentation of the building footprint. Figure 50 and Figure 54 illustrate the results of the simplification for three example buildings.

A Property Boundary Cutting through two Connected Buildings

With the elevation values along the footprints, we have the 3D outlines of the buildings. At issue is the situation along a parcel boundary where there may be a valid building footprint, or the building is attached to a structure on the adjacent parcel and the footprint is merely virtual. To determine whether the footprint is virtual or real, we revisit the elevation data. Along a footprint at the edge of a parcel, one defines a small mask of perhaps 20*20 pixels. If one is dealing with a real footprint, then half of the elevation values should be zero. If the footprint is virtual, then a majority of the elevation values will be large. We select a threshold of 2/3 of all values to be large enough to determine that the footprint is virtual. Figure 55 shows such a façade with 2 virtual facades.

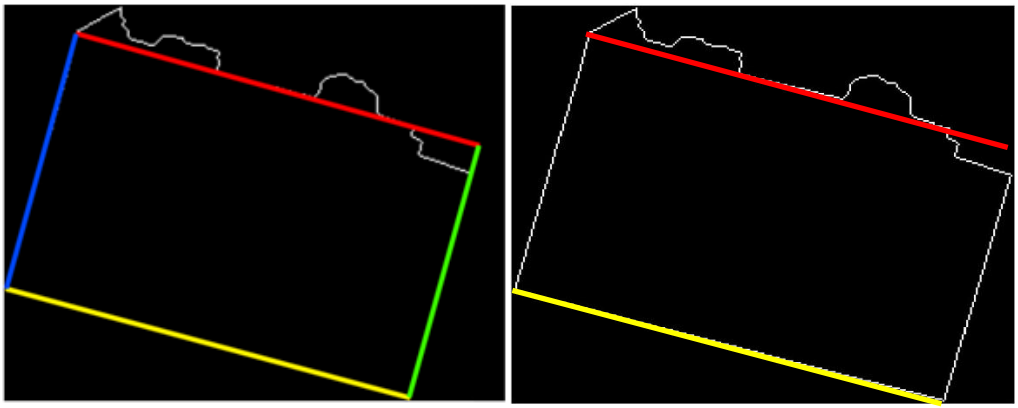


Figure 55: Building Footprint with two virtual facades (highlighted in blue and green) and footprint corrected footprint

Determination of Facade Quadrilaterals

Another important initial parameter for every building valuation is the height of the single buildings. Therefore, we have to employ the height field (DSM) and isolate the buildings, separate from features that do not belong to the building (cars, trees). This feature of object removal is just working if high redundant image information is used.

The use of the DTM in defining footprints produces, as a byproduct, an estimate of an elevation value for each candidate footprint. While this has been computed for candidate positions where a footprint location is possible, this now needs to be converted to a set of elevation values along the path of the footprint. For this purpose, the geometric figure of the footprint is placed into the DSM and the elevation profiles get interpolated along the straight lines of the footprint: For each straight line of the footprint repeat the following process:

- Define positions i,j along the straight footprint line at equal intervals;
- Determine the XY-pixel locations perpendicular to the line at positions i,j;
- From the short elevation profiles along the pixel locations XY, determine the base height and the top height associated with that footprint element, and thus the elevation difference.

The result of this procedure is a set of elevation profiles along the footprints. Depending on the shape of the roof, a façade could also have a complex shape. Should the elevations change along the façade footprint, we could be facing a sloping ground or a varying roof line, or a combination of both. The slope of the ground is known from the DTM. Variations of the roof line are available off the difference between the DSM and the DTM and must be considered.

Decomposing Complex Buildings into Separate Building Objects

For the determination of the complexity of a building one has to review the roof shapes. Having more than one local maximum in the roof height is an indication that the single building has a complex shape and should be segmented into multiple building objects/ sub-buildings.

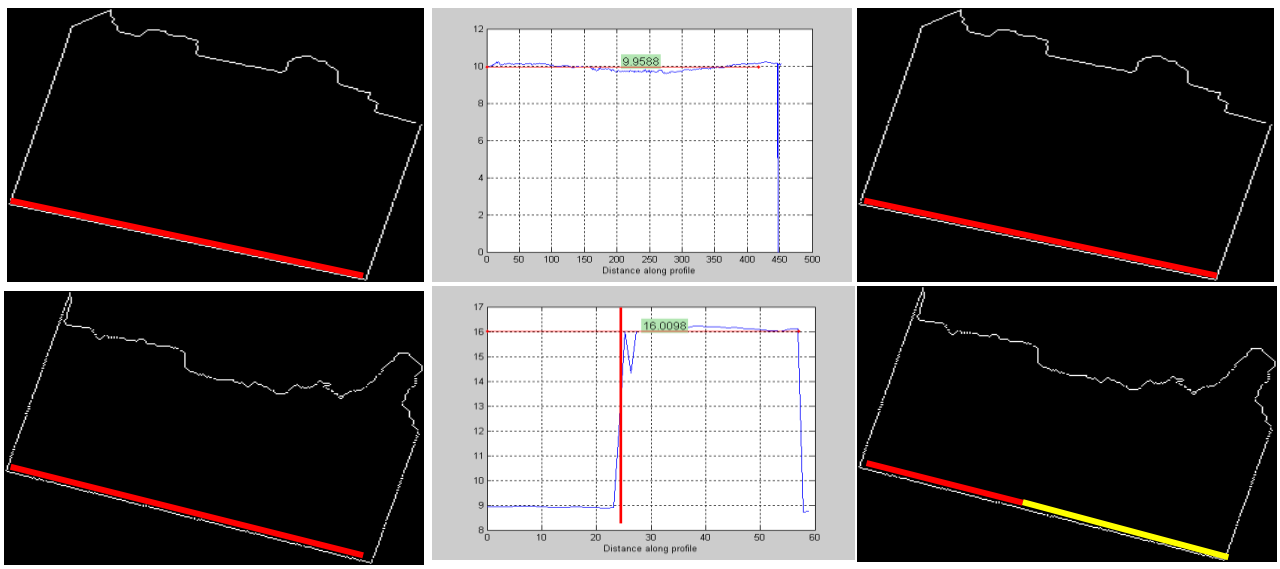


Figure 56: Images shows 2D footprints and height profile along these footprints of two buildings (above) Simple footprint with constant height profile. (below) Complex footprint with height discontinuities along elevation profile (highlighted in red in the histogram); Red line illustrates façade split.

The split of a complex building into simpler building elements has been discussed by [Zebedin et al., 2008] and implemented in a workflow to replace a dense point cloud by simple building geometries. There exist three measures of complexity for a building. One is the geometric figure of the building's footprint. One may restrict the complexity of a footprint as discussed earlier. The second is the elevation profile along the footprint. One may determine a measure of the building symmetry for the elevations along the footprint: if facades get associated with different building heights, one may have reason to break the building into its parts (see Figure 56). The third is the number of local maxima in the elevations of the roof: the roof shape is defined by the elevation values inside the footprint figure. By computing local maxima for those elevations, one will have the means to determine a separation of the building into building elements, each with a separate roof. [Zebedin et al., 2008] evaluates the height differences between manual and automatic reconstruction of a building for a test data set of Manhattan (1973 buildings). It shows

that 67.51% of the pixels have a height difference smaller than 0.5m, 72.85% differ by less than 1m and 86.91% are within 2m. Details on this method are described in [Zebedin et al., 2008].

Small Structures versus Buildings

With elevation profiles along the footprints, we also have the means to separate actual residential housing from detached garages. The latter will have a small surface area of 50 m2 or less and not exceed a height value of 2.5 m.

Projecting the Façade Quadrilaterals into the Aerial Images

The façade's geometry is assumed to be a quadrilateral and has been produced from the intersection of the façade footprint with the DSM. This delivers 4 façade corner points in 3D object coordinates that get projected into the aerial imagery. Therefore, the three dimensional object points are mapped to a two dimensional image plane using a central projection.

6.3. Experiments

We have performed the building façade detection for the Graz test area with 216 buildings on 321 parcels. In a preprocessing step, we eliminated small structures that are classified as buildings but don't fulfill certain criteria. In our test area, 54 small structures were eliminated. In a first step we determined 2847 straight line segments based on the building classification and 2895 line segments based on the DSM. This result shows that the building classification and the DSM produced nearly the same segments. This is due to the fact that the building classification is related to the DSM. The following merging step by calculating the mean between the two results produced 2855 line segments. To enable a fast characterization of the façade footprints, a simplification of these line segments was performed depending on their orientation, neighborhood relationships and length. We reduced the number by a factor of four to 725 simplified line segments. In a next step, all 262 virtual facades are eliminated resulting in 463 remaining building façade footprints. By incorporating the elevation values we determined 20 complex building facades with two or more local height maxima. By applying the collinearity equations we transformed all four corner points of each façade quadrilateral into the corresponding aerial images.

Figure 57 shows a true orthophoto of our test area in Graz, and a particular building marked in red. In the lower right corner of the true orthophoto, the relevant building footprint is shown with one façade highlighted in red. The four corner points are projected into all relevant aerial images. A visual inspection shows that in all three cases the façades could be successfully extracted.

These 463 façade quadrilaterals were projected into the corresponding vertical aerial images. In our case, the test area is covered by 20 vertical aerial images. With a forward overlap of 80% and a side overlap of 60%, the façade quadrilaterals are visible in up to 15 images that results in a total of 5093 façade images.

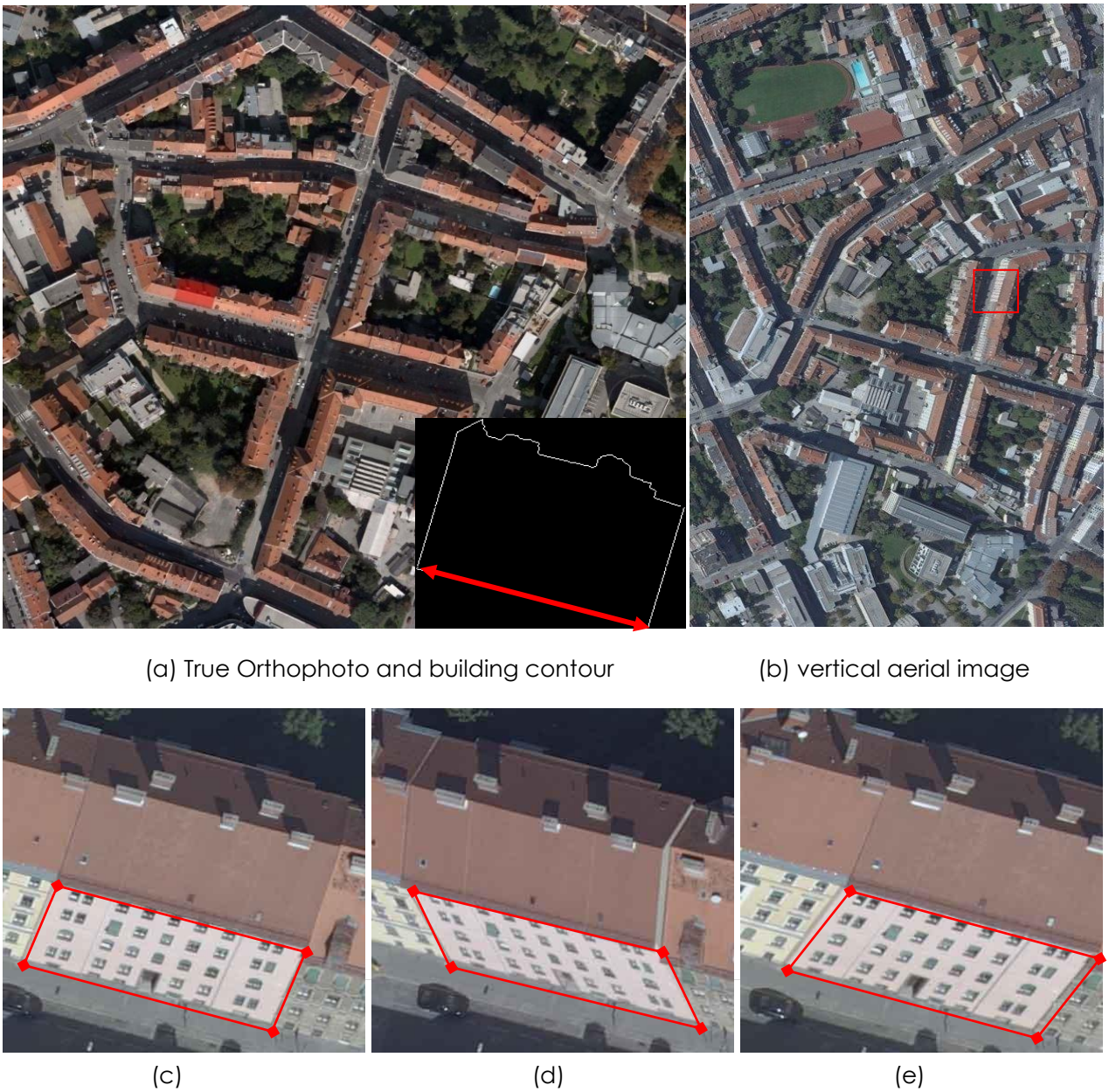


Figure 57: (a) True orthophoto of Graz test area with example building highlighted in red and relevant façade footprint; (b) one vertical aerial image showing the respective building façade. (c-e) Corner points of one façade of one building projected in three different vertical aerial photographs. The façade quadrilateral is superimposed over the aerial photography to obtain the photo texture for visual inspection.

CHAPTER SEVEN

FAÇADE DETAIL CHARACTERIZATION

Contents

7.1	Introduction	145
7.2	Hidden Façade Removal	146
7.3	Planar Facades	147
7.4	Complex Facades	151
7.5	Experiments	160
7.6	Conclusion	169

7.1. Introduction

We proceed from the basic building description to an analysis of the facades. A characterization of building facades presupposes that we have a good understanding of its definition. We use a definition for facades found in Wikipedia:

"*A facade is generally one side of the exterior of a building, especially the front, but also sometimes the sides and rear.*"[27]

A façade is generally assumed to be a plane, but with superstructures, much like a roof is assumed to be a plane, but with many extrusions. Work on building facades would intuitively build on street level photography or at least oblique aerial imagery. However, our interest is in understanding how well one can study facades in (cost free) vertical aerial photography. The façade detail of interest is a floor count, a window count and a description of the window's location and size.

To address a count of floors and windows, we need to cope with both truly planar as well as complex facades. We compare the façade footprint with the building contour. If the distance values lie within a certain threshold, in our case ±50 cm, a facade is considered to be planar. If not, the facade is considered to be complex (see Figure 58). This classification is crucial for the

[27][http://en.wikipedia.org/wiki/Façade]

further characterization of its details. Depending on this classification, different processing steps are necessary to achieve meaningful results. We split this chapter into two parts. In the first part, we deal with planar facades where we also explain the proposed method for floor and window detection. In the second part, we discuss complex facades and how we have to handle them to be able to characterize them. However, before we can characterize these facades we have to check their visibility.

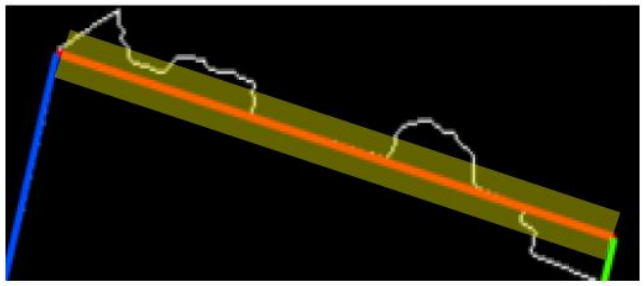


Figure 58: Determination of planar and complex facades (white line shows building footprint derived from building classification; red line is simplified façade footprint; yellow area contains points that lie within 50cm of the footprint.)

7.2. Hidden Façade Removal

The determination of the façade quadrilaterals and their projection into the different aerial images helps to determine their visibility. Depending on the location of the camera center, facades are visible or hidden by the building (see Figure 59). To determine if a façade is visible we use a method widely used in Computer Graphics called back-face culling [Cormen et al., 1990, Luebke et al., 1995]. It is a method to eliminate the backfaces of on object if they are occluded by the object. Therefore we compute the intersection between the two diagonals of one façade M and calculate the normal vector N. If the normal vector shows away from the camera location, the façade image is eliminated. Therefore we also compute the vector between the camera location P and the intersection M of the two façade diagonals and calculate the dot-product of the two vectors $V = \overrightarrow{MN} \circ \overrightarrow{PM}$. Depending on the result we can eliminate the false façade images.

Figure 59: Building facades extracted from different overlapping aerial images; (above) three façades that are visible in aerial images and thus can be further processed (same as in figure 60) (viewing angles: 25°, 26.2°, 19.7°) and (below) two images where façades are hidden and thus have to be eliminated.

7.3. Planar Facades

The computed visible planar façades can now be used for further processing. Before we deal with the floor and window detection in a preprocessing step, we have to rectify them and perform a contrast enhancement. The rectification of the image content is applied by using a 2D bilinear interpolation to resample the image in a novel view. Typically, overlaps cause many aerial images to show the texture of a given facade. Figure 60 presents one façade of a building from the Graz test data set shown in four different vertical aerial images after rectification. Although the resolution of these facade images is very coarse due to the steep viewing angles, the image information is sufficient for floor and window detection.

Figure 60: One building façade visible in four different vertical aerial images. These have been rectified into a façade coordinate system. From an aerial image block showing for each object point typically 10 images, not all will contain useful data for a specific vertical façade. Selected here are the 4 best, where "best" is defined as the largest area of a façade quadrilateral in the projection into an image.

Number of Floors

The determination of the number of floors results from the building height and the average floor height. Because this height is often dependent on the age of the building, it is necessary to get this information from the title register or from other databases.

Another possibility, without using the title register, to determine the date of origin of a building is from the roof form. Here we can assume from the fact that a flat roof cannot be found on older buildings, that all buildings with that type of roof are likely to be newer ones. That means that floor heights of > 3 m are not usual (average height 2.5m). By contrast, buildings with steep gable roofs are most likely to be old buildings that have higher rooms (on an average 3.8m). The thickness of the covers must be considered within this calculation as well (on an average 50cm) and must be added to the floor height. However, this method is error prone because not all buildings have been built in this way.

Another more promising method is the incorporation of windows on a façade. From the building's appearance, floors normally get defined by windows. Where this fails may be in industrial plants or some contemporary public architecture. In any event, windows will always be a defining element of any façade, and will also be associated with floors. A procedure for finding a floor count has been developed using a projection profile based approach. We use the fact that building windows are horizontally or vertically aligned with a regular spacing. It addresses a façade in a single plane and without significant extrusions. The proposed floor detection approach of Table 22 is illustrated in

Figure 61. A floor count can be applied to each set of overlapping façade images. If there were a discrepancy in the result, some logic would have to be applied to resolve the ambiguity. Another limitation may exist when the edge histogram fails to have a distinct extremum. This might indicate that the façade/building is complex and that a refined approach should be used. One such option is the development of a 3D façade point cloud, as discussed in a later section.

For each façade i of a building j, repeat:

- Import all n image segments showing this façade i.

For each image segment repeat:

- Transform the segment into the façade coordinate system.
- Apply a contrast enhancement.
- Apply the edge detection horizontally.
- Apply the edge detection vertically.
- Convert the maximum horizontal and vertical edge values into a binary format.
- Projection of each image row, and image column in a vertical and horizontal edge projection profile.
- From the summation, remove outliers, normalize the values and remove low values as "noise".
- Determine the number of maxima of the sums of vertical gradients and use this as the number of floors.
- Perform a verification by eliminating floors that do not have the proper vertical spacing (minimum distance between floors); and remove values from along the edges of the image texture inside the façade quadrilateral.

Table 22: Procedure for floor detection in simple planar facades

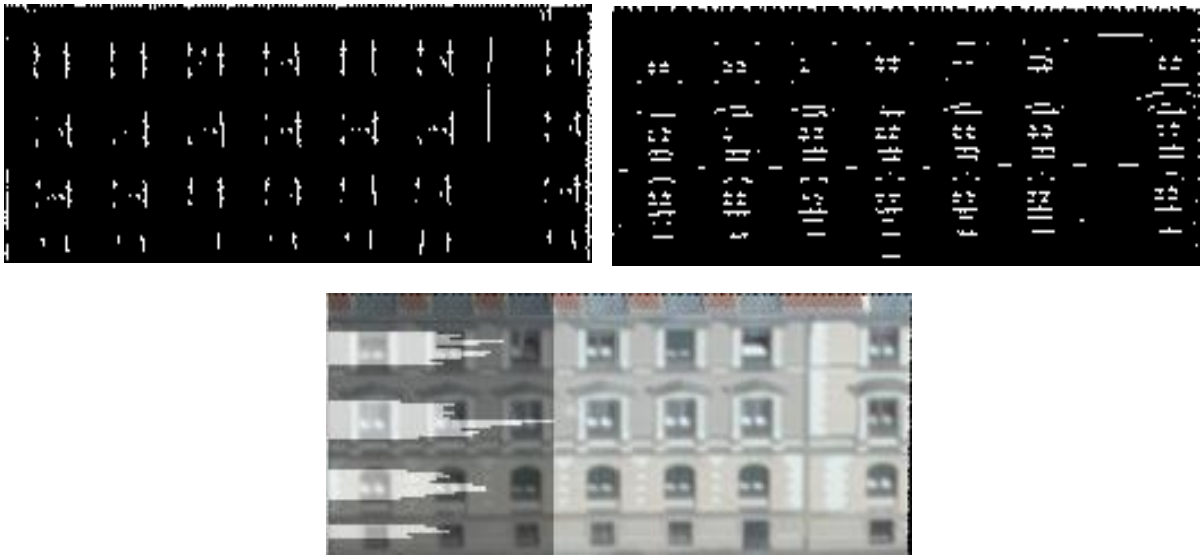


Figure 61: Binary Prewitt edges in (a) are vertical, in (b) horizontal. The sums of edge values are shown in (c) as a count of the number of floors

Furthermore, we could develop additional information from every image about the building that is important for the valuation such as:

- Is there a garage in the building?
- Is there a cellar? (if windows of the cellar are visible in the images)
- Next, a verification on whether or not an attic exists has to be done. This can be done from the style of the roof and the roof structure. From aerial images and oblique images it is possible to detect dormers and bigger windows on a roof. If there is no extension of the attic, this information is also stored and used in the valuation (possible increase of the building value).

However, these are secondary issues in the current research plan. While a determination of the number of floors is essential, and is a primary factor of "value", those secondary parameters will be assessed as the opportunity exists.

After the determination of the number of floors, the floor-space index is available by multiplying the number of floors with the building base. Masonry and staircase represent, as a rule, about 10% of the base. Thus, on has to subtract the masonry and the staircase of the building from the base of the house to receive the floor-space index.

Windows

Window detection has been of some interest in recent years. Algorithms like "boosting" have been applied by [Grabner et al., 2007] to detect cars in aerial photography and windows in street-level images. [Cech et al., 2007] have developed a window detection based on a library of window shapes, analyzing street-level photography, as did [Lee et al., 2004] who based their approach on edge images. These approaches have been subjected to only limited experimental analysis, but are generally reported to find windows in a rather robust manner. In the current application, we need to deal with vertical aerial photography and source data with more noise than street level images have.

Given floor counts, we are reusing the intermediate detected horizontal and vertical edges to also find the windows per floor. An approach that simply "intersects" the locations along the pixel rows and columns with the maximum edge sums will work if all windows are regularly arranged. The simple intersection of the vertical and horizontal edges may not be highly accurate, especially when there exist different shapes of windows in one column. Therefore, [Lee et al., 2004] have proposed a variation of the approach. To refine the locations of the windows a one dimensional search for the four sides of a window is performed. Hypothesized lines are generated by moving the line to its perpendicular direction. The hypothesized line with the best score determines the position of the window boundary line. We find this method applicable to images with lower resolution, as is the case in facades imaged in vertical aerial photography. Also, not only rectangular windows but other window designs can be automatically and quickly detected without training the program in advance. The window count is applicable in each image segment of a given façade, separately. Or one might want to merge the edge data sets and apply single window detection to the sum of all edges.

We may get different floor numbers from the multiple facades of one building. It is possible that a building has different floor counts on a sloping terrain. Since the "bald Earth" as well as the slope of a building footprint is known, this must enter into the floor count. Also, we have redundant information from overlapping images all showing one and the same façade. Figure 62 illustrates four results for one single façade imaged four times. In this example, all 28 windows of the façade were found in all 4 aerial images. The result of thesis processing is the sizes and the quantity of the windows per floor which can be considered in the valuation.

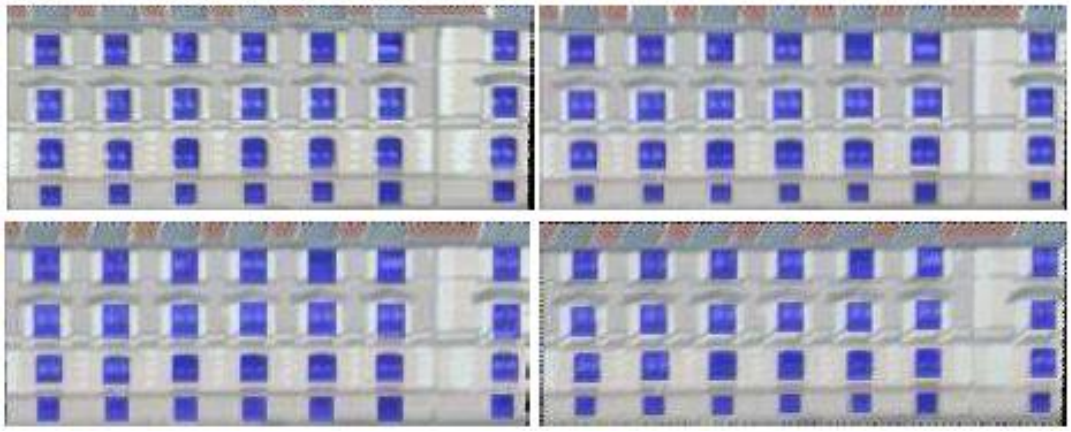


Figure 62: Four images on one façade from Figure 60 lead to independent floor and window counts. It is to be noted that the floor counts and the number of windows (highlighted in blue) coincide with the visual inspection (see figure 63).

7.4. Complex Facades

Facades are not always in a single plane. Figure 63 presents a typical example of a building side consisting of various details such as extruding staircases and awnings. The application of the previous workflow to this type of façade will fail because there will not be usable edge histograms, and the process will deliver a failed classification. A possible elimination of these problems could be a splitting of the façades into multiple façade fragments. A more promising solution is the explicit consideration of the 3rd dimension that generates 3D façade point clouds.

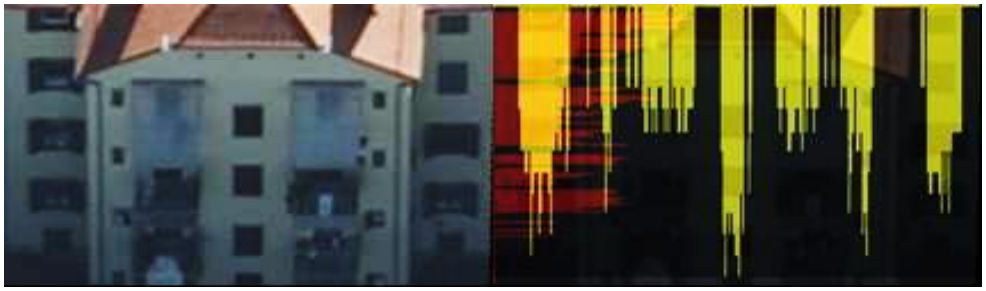


Figure 63: To the left is a rectified facade image with its depth structure. To the right is a failed count of windows using the workflow that assumes that a facade is in a single plane. Abandoning this restriction will cause a need for 3D point clouds of facades.

One will have to examine whether 3D façade point clouds can be produced at a sufficient detail and resolution from vertical aerial imagery, so that the facades can be successfully analyzed. "Success" will be achieved if the 3D point cloud (a) improves the segmentation of a façade into façade planes, (b) then produces well defined values for the floor and window counts on these façade parts, and (c) finally supports the reconciliation of the façade parts into a single window and floor count for that building. To be able to do this the following processing steps have to be executed (see Figure 64):

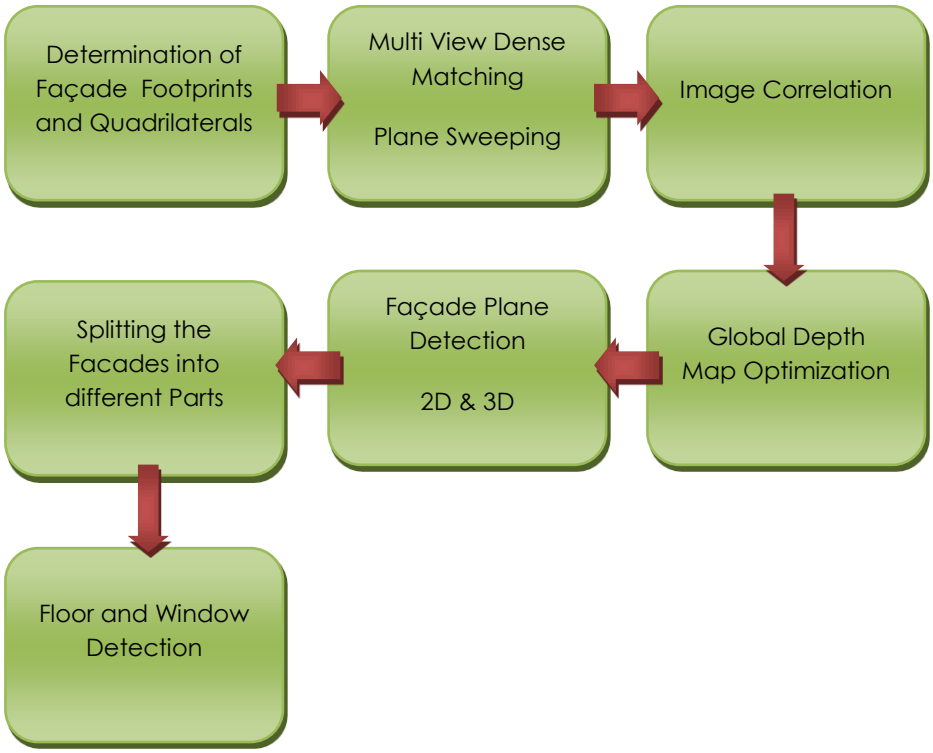


Figure 64: Roof Characterization Framework

Multi View Dense Matching

Performing a photogrammetric 3D process on a set of overlapping images of a facade taken from the street level is a widely studied endeavor [Dulong et al., 2010]. The challenge in the current context is different since we rely on vertical aerial photography where facades suffer from major geometric deformations and non-square pixels. The approach is the same as that used for DSM-creation from highly redundant photography [Irschara et al., 2007]. The workflow has been described by many computer vision laboratories, for example [Gösele et al., 2010]. At its core is a so-called plane sweep [Baillard et al., 2000; Collins, 1996; Gösele et al., 2010]. The plane sweep approach is a well established method in computer vision for an image based reconstruction from multiple views. This is in contrast to traditional computational stereo methods. A 2D space is defined by multiple planes that lie parallel to the key-plane (see

Figure 66). A key-plane, the approximated façade-plane, is determined separately for every building façade. This is done in the context of the determination of the façade footprints. In addition for every key-plane we have to select a key-view. Therefore we use two constraints: (a)

the angle-of-view α and (b) the direction of the façade in dependency on the camera location β (see Figure 65).

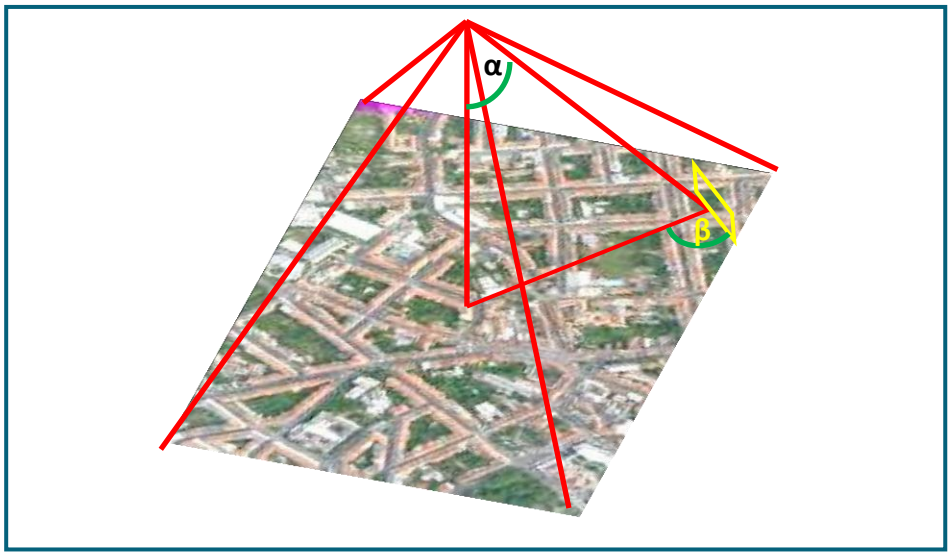


Figure 65: Definition of the two constraints angle-of-view α and direction of the façade β.

In a next step, we generate multiple planes parallel to every key-plane. The plane at a certain depth from the key view induces homographies for all other views, thus the sensor images can be mapped onto the current plane $\pi = (n^T, d)$.

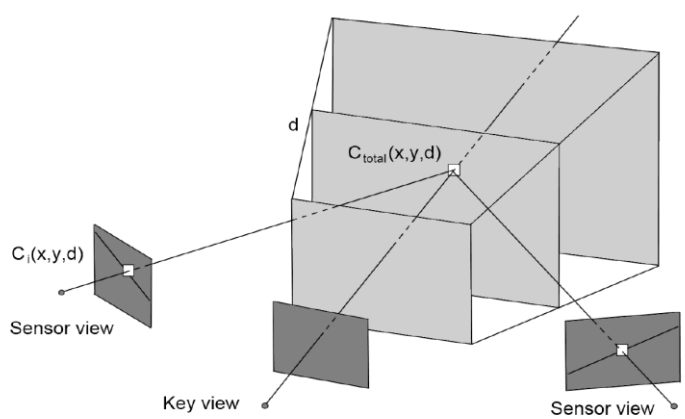


Figure 66: Plane sweeping principle. The homography between the façade's reference plane and the sensor view varies for different depths. [Szeliski et al., 1999]

The sensor images are warped onto the current 3D key plane using the projective transformation. The key view is given in canonical coordinates $P = K[I|0]$ according to the appropriate homography H with

$$H = K\left(R - \frac{tn^T}{d}\right)K^{-1}$$

where n is the normal vector of the plane and d is the current depth hypothesis (distance between planes). K is the intrinsic matrix of the key view and R,t is the relative pose of the sensor view $P' = K'[R|t]$ with respect to the key view. The relative pose can be computed from

$$R = R_2 * R_1^T$$

$$t = t_2 - R_2 R_1^T t_1$$

Derived from the two projection matrices $P_1 = K_1[R_1|t_1]$ and $P_2 = K_2[R_2|t_2]$ and the normal vector plane n = [0,0,-1].

Image-space algorithms usually constrain the maximum disparity range or interval, in which depth values can occur. Respectively, the extent of scene geometry is determined to lie between a near and far plane from the camera center of a key view. The minimal and maximal depth range depends on the complexity of the respective façade. Its complexity is determined by overlaying the key-plane with the façade footprint and computing a distance image (see Figure 67). This image is then used to calculate the appropriate number of depth steps to either sides of the 2D façade footprint. The depth sampling Δd and the number of depth steps is chosen so that sub-pixel accurate matching is achieved. The depth steps can be calculated so that the Nyquist criterion [Shannon, 1949] f > 2 pixel is satisfied for at least half of the sensor views.

$$median\left(\left\|p\left(P_i, X(d)\right) - p\left(P_i, X(d + \Delta d)\right)\right\|\right) < 0.5\, pixel$$

X(d) is the point passing through the center of every tile at depth d, P_i is the projection matrix of sensor view i and p is the projection operator.

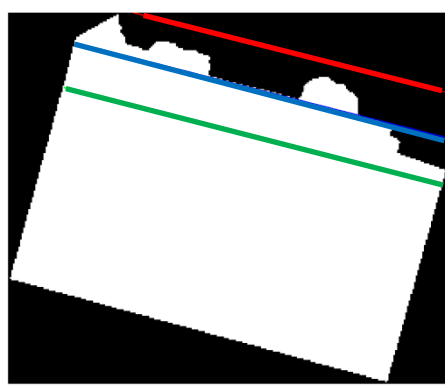


Figure 67: Minimal and maximal depth range for one complex façade (blue: approximated façade footprint; green: maximal depth range; red: minimal depth range);

Image Correlation

After computing all parallel planes and corresponding sensor views, we then overlay the key-view with all sensor views of one plane and calculate the cross correlation coefficient. If the plane at a certain depth passes exactly through parts of the object's surface to be reconstructed, a match will exist between the relevant parts of the new sensor view and the key view. By sweeping the plane through the 3D space, a cost volume is filled with image correlation values corresponding to the disparity space image (DSI) in traditional stereo [Seitz et al., 2006]. After projecting a sensor image onto the current plane hypothesis, a correlation score for the current sensor view is calculated. The final correlation score of the current plane hypothesis is achieved by integrating all overlapping sensor views. For the accumulation of the single image correlation scores, a simple additive blending operation is used. We use zero-mean normalized cross correlation (ZNCC) as photo consistency measure. The correlation between two images (cross-correlation) is a robust approach for dense matching. One of the advantages of ZNCC is that the correlation coefficient r for window-based local matching of two images X and Y is invariant under affine linear changes of luminance between images. To receive robust results for the determination of the correlation coefficient, neighboring pixels are added up. This can be done for a neighborhood of e.g. 3x3 pixels. Given two intensities X and Y, the ZNCC is computed by

$$r = \frac{\sum_{i \epsilon W}(X_i - \tilde{X}) * (Y_i - \tilde{Y})}{\sqrt{\sum_{i \epsilon W}(X_i - \tilde{X})^2 * \sum_{i \epsilon W}(Y_i - \tilde{Y})^2}}$$

Where $\tilde{X}, \tilde{Y}$ denote the mean intensities inside the rectangular region W. We repeat this process for all parallel planes and all corresponding sensor. The results of this calculation are n_k matching probabilities for every pixel x(i,j) of a façade for all n façade planes. Figure 68 shows the result of this correlation for 4 different planes.

Figure 68: Correlation coefficients calculated for 4 different planes visualized as binary images (white areas have the largest correlation values).

Depthmap Generation

For the determination of the depthmap we use a total generalized variation multi-labeling approach proposed by [Zach, 2007; Pock et al., 2008]. Because of poorly textured areas and repetitive structures in the façade images we receive unstable results. To avoid these problems

we use the 3D structures spatial coherence and employ various surface views in a multi-image approach. From the 3D cost volume, dense depth maps can be extracted using global optimization methods. Given a graph with node set V, edges E and a label set $L \subset Z$, an optimal labeling $l \in L^V$ for the energy of the form,

$$\min_{l} \sum_{(u,v)\in E} P\big(l(u) - l(v)\big) + \sum_{v\in V} D\big(l(v)\big)$$

$P\big(l(u) - l(v)\big)$ are pairwise potentials and $D\big(l(v)\big)$is the unary term, respectively. Solving this problem corresponds to a minimal cut on a graph in higher dimensions where labels are ordered. In [Ishikawa, 2003], a minimum cut algorithm is presented that exactly solves this class of Markov Random Field (MRF) problem. This problem perfectly fits to dense depth estimation, where $l(v) \in L$ are depth labels, $v \in V$ pixels and E describes the connection of pixels. Such a labeling combines a certain pairwise regularity term P(.) with an arbitrary data term D(.). In [Pock et al., 2008], a continuous formulation to the discrete multi-label problem of Ishikawa is given. The corresponding variational problem is,

$$\min_{u} \left\{ \int_{\Omega} |\nabla u|\, dx + \delta \int_{\Omega} g\big(x, u(x)\big) dx \right\}$$

Where$\Omega \subseteq R^2$ is the image domain, g(x,u) is the image matching term for each surface point and δ is a penalization value. The energy consists of two terms, a smoothness term to model the spatial coherence of the surface and a data term that reflects the multiple image matching qualities. The left term is the so called total variation (TV) and is used for regularization, ensuring that the façade surface is smooth and preserving sharp discontinuities in the solution while remaining convex. It is based on minimizing the TV of u. The data term to the right measures the matching quality. These results can directly be converted to a 3D depthmap and visualized (see Figure 69). For a better evaluation of the results, we have computed 3D point clouds using different values for δ. By varying the value used for δ we can influence the level of complexity of one façade. Depending on that, the calculation preserves smaller structures or not. This is very important for instances of the detection of balconies or bay windows (see Figure 69).

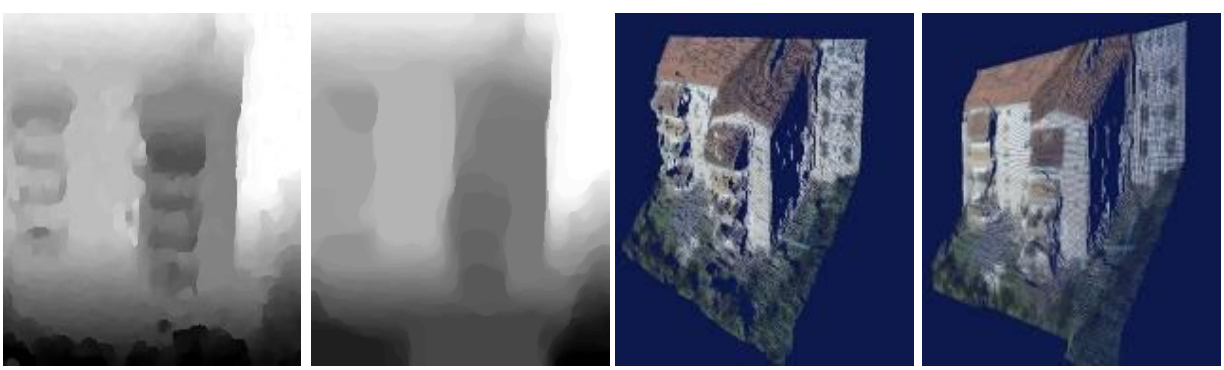


Figure 69: (a) Computed depthmap of one façade (see figure 73a) and (b) overlaid with photo texture (left: δ = 10) (right: δ = 2)

The spatial continuous formulation contains several advantages over the discrete approach. On the one hand, continuous optimization can be implemented using simple and efficient primal-dual optimization techniques which can be easily accelerated on parallel architectures such as graphics processing units (GPUs). On the other hand, these methods require considerably less memory which makes the method applicable for quite large practical problems [Pock et al., 2008].

Major Facade Plane Detection

Applying the information from the 3D point cloud we now can refine our definition of planar and complex facades. We determine if a façade is planar by calculating a regression plane of the point cloud, and by determining how many points of that point cloud are part of the regression plane. This determination depends on two constraints, (1) how many points lie within a certain range of the regression plane and (2) how far are the other points away from this plane.

A further aspect when characterizing complex facades is the classification of the different overhangs. In the scope of this work we detect 3 classes: (a) staircases and bigger parts of a building, (b) bay windows and (c) balconies.

The difference between bay windows and balconies lies in the homogeneity of the depth values. Balconies usually have more diverse depth values than bay windows. Bay windows are smaller parts of a façade with homogenous depth values.

Next, one has to analyze the 3D points with the goal of segmenting the facade into its planar sub-facades. For plane detection in 3D point clouds various methods are available. There are two main possibilities to characterize 3D point clouds: Determination of façade planes using 2D or 3D.

2D interpretation of the point cloud

We project the 3D point cloud in both the xy-plane and yz-plane and then characterize the resulting profiles. The advantage of this method is that it works fast and is easy to implement. However, we reduce our 3D point cloud from 3D to 2.5D because we use the most common depth values for every façade.

We start by eliminating all outliers for every column and search for local maxima in the single columns. Then these values are weighted according to their appearance in the column. In a next step, we look for maxima in the xy- plane (footprint of façade) (see Figure 70). These maxima determined in the xy-plane are the major façade planes of one facade. After that we detect regions where balconies are located. This is achieved by looking at the projected footprint of the façade and searching for rough areas (highlighted in green). After we have detected our major façade planes, we can project the major planes in the yz-plane to determine if there are roof overhangs. Therefore, we use the same processing steps as before. We eliminate all outliers for every row, look for local maxima in every row and weight their appearance. The result of this calculation is shown in Figure 70c.

An advantage of this method is that we can use the computed footprint (xy-plane) and profile (yz-plane) to enhance our building classification and to determine the emerging masonry

of one façade, which is usually not possible when dealing with vertical aerial images (see Figure 71).

(a) (b) (c)

Figure 70: Plane detection for a complex facade; (a) key-view of a facade and below projection in xy-plane (red lines are strongly weighted (>66% of points lie in this plane and green lines are balconies); (b) determined major façade planes (segmented areas with balconies are marked in red); (c) side-view (yz-plane) of turquoise façade plane (projection of points into yz-plane)

The major disadvantage of this method is that it neglects the availability of 3D and therefore smaller building overhangs like bay windows are ignored. Also, if facades have a more complex structure the 2D approach has major problems. To avoid this problem and to cope with more complex buildings we also use a 3D interpretation of the point cloud.

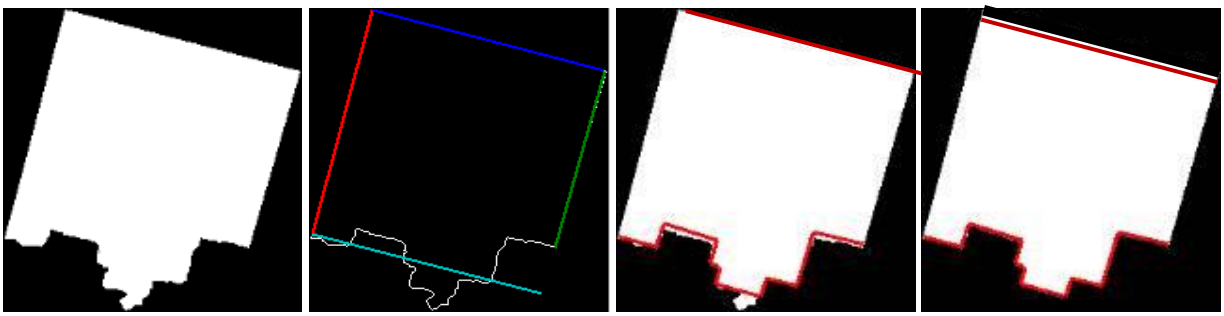


Figure 71: Enhancement of building classification. (a) Building Footprint of one building; (b) determined façade footprints; (c) overlaid enhanced building footprint (d) modified enhanced building footprint.

3D interpretation of the point cloud

The most common method in computer vision is the classical RANSAC plane detection [Fischler et al., 1981; Förstner et al., 2010], which we discuss in more detail in Chapter 8. It selects randomly 3 points in the point cloud, calculates the parameters of this plane and assigns, according to a given threshold, all 3D points that belong to this plane from the 3D point cloud.

This procedure is repeated n-times and the obtained results are compared with the last saved one. At the end, the best fitting plane will be computed. In our case, the determination of the major façade planes using RANSAC or similar algorithms is not necessary. The reason is that we have already assigned probabilities to every depth value of the 3D point cloud in the scope of the multi-view dense matching procedure. So what we have to do is use these probabilities and assign them to façade planes.

In this case, for the labeling of the major planes, the following processing steps are necessary:

- Elimination of outliers by looking at the depth value distributions
- Identification of approximated major planes by looking for local maxima (see Figure 72)
- Calculation of distances between single depth values and all approximated major planes
  Labeling of the major planes using Potts model (see Figure 72)

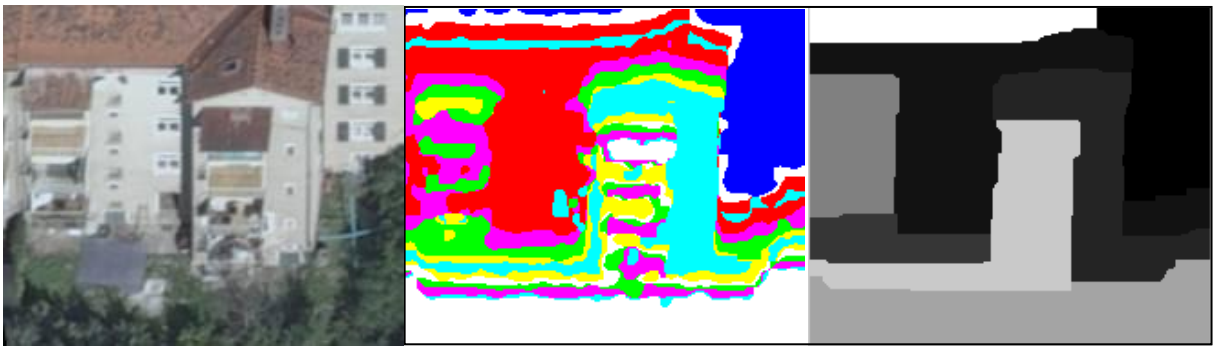


Figure 72: Plane detection using 3D approach; (a) key-view of facade; (b) approximated major façade planes without regularization; (c) labeled façade image

The advantage of this method is that the 3D character of a facade is preserved (see Figure 72). Additionally, using the 3D point cloud allows a segmentation of a complex facade into balconies, bay windows and other facade parts. Figure 72 shows two areas where balconies were determined.

Façade Plane Interpretation

In a next step, we split our façade into parts according to the results of the point cloud interpretation (see Figure 72). The detected façade parts can now be cut in the corresponding sensor-views and the key-view. This is possible because the segmented façades include information about the respective planes. In a refinement step, small and narrow regions and regions next to the upper and lower border of an image are excluded. The determination of regions nearby the borders works by calculating the position of the centroids of all regions in an image. If the centroid lies within a certain distance of the upper and lower border the regions are identified and excluded from processing. Also, horizontal narrow regions are excluded if it is a bay window. For the remaining regions a floor and window detection is performed. Figure 73 presents the result of floor and window detection. If one compares these results, of the window and floor detection, with the previous results from Figure 63 one can see the improvement.

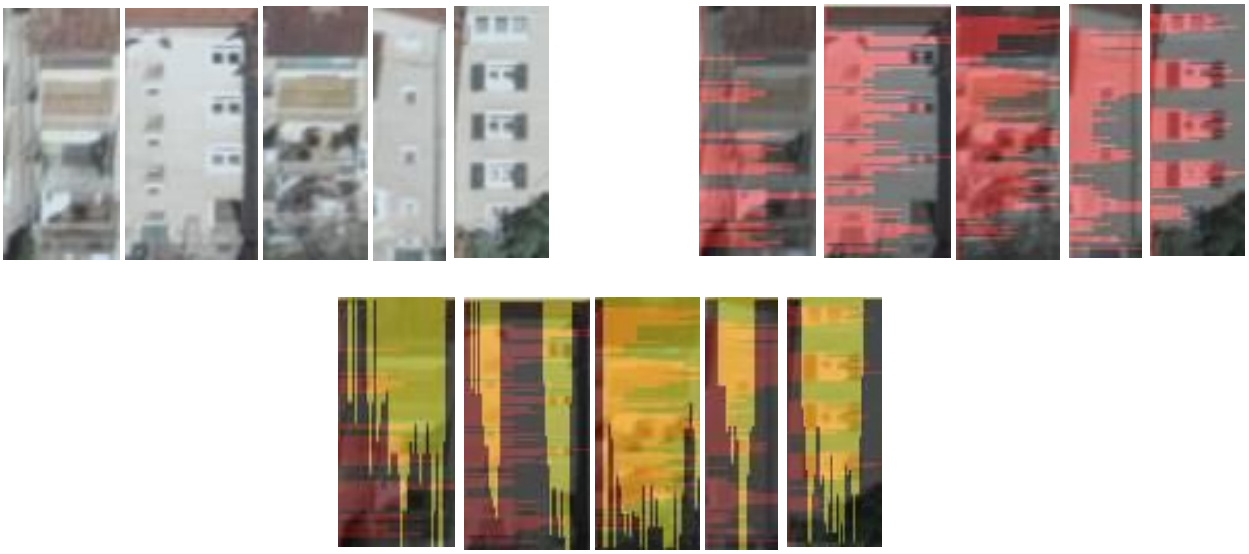


Figure 73: Floor and Window detection for the determined labeled regions using the key view

7.5. Experiments

To evaluate the characterization of planar and complex building facades we have performed experiments in the Graz test area. For this purpose we use 216 buildings with a total of 463 façades. Before we can characterize these façade images we perform two preprocessing steps. In a first step, we eliminate all hidden facades using the back-face culling principle. In a second step, all facades are eliminated that are photographed below a certain angle-of-view. After these preprocessing steps, we perform a floor and window detection for planar facades and for complex facades.

Hidden facades and elimination of "bad" façades

We perform back-face culling to determine the visibility of building facades. The façade quadrilateral determination could detect for the Graz data set a total of 463 building facades visible in a total of 5093 aerial images. After the visibility check, 2287 facades images were detected as hidden facades resulting in 2806 remaining façade images for further processing. In a next step, all façade images with insufficient viewing angles were eliminated. In our case, we defined the threshold as follows: viewing angle α <10 degrees or if their orientation β is outside a certain range (< 45 respectively > 135 degrees) (see Figure 65); This processing step eliminated additional 1354 façade images resulting in 1452 images for the façade characterization. Table 23 summarizes the classification of the façade images into different viewing angle categories.

Angle α [deg]	< 5	5 - 10	10 - 15	15-20	20-25	> 25
Façade images	354	407	497	501	539	508
Façade images	0	0	327	344	409	372

Table 23: Façade images classified into different viewing angle categories before incorporating rejection thresholds and results after incorporating rejection thresholds.

As illustrated in Table 23 only 761 facades were eliminated by considering the viewing angle α. 593 façade images were additionally eliminated by applying the orientation threshold.

In Table 24 we have categorized the facades into different classes depending on the number of images that show these facades. As one can see, most of the facades are found in three to four images. Unfortunately, because of the constraints concerning angles, there are 17 facades that are visible just in one image.

	1 image	2 images	3 images	4 images	5 images	total
No. Façades	17	127	145	104	66	463
No. Images	17	254	435	416	330	1452

Table 24: Classification of the building facades into different categories depending on the number of images that are visible.

In a next step, we classify the facades into two categories: planar facades and complex facades. Therefore, we compare the façade footprint with the building contour. If the distance values lie within a certain threshold, in our case ±50 cm, a facade is considered to be planar. If not the facade is considered to be complex (see Figure 58). Table 25 summarizes the result of this classification. In the Graz data set, 286 facades are planar visible in 897 images and 177 are classified as complex visible in 555 aerial images.

	planar	complex	total
Façades	286	177	463
Images	897	555	1452

Table 25: Division of the facades into complex and planar facades

Analysis of Planar Façades

In the next section, we want to evaluate the floor and window detection for the Graz data set consisting of 216 buildings with 286 planar facades. For the evaluation we classify the planar façades into different viewing angles and visibility classes. Table 26 summarizes the classification

of the planar facades into the different viewing angle classes. Table 27 illustrates an evaluation of these planar façades concerning visibility in the aerial images.

Angle α [deg]	*< 5*	*5 - 10*	*10 - 15*	*15-20*	*20-25*	*> 25*
Façade images	0	0	189	178	260	244

Table 26: Planar Façades classified into different viewing angle categories.

	1 image	*2 images*	*3 images*	*4 images*	*5 images*	*total*
No. Façades	10	81	90	70	35	286
No. Images	10	162	270	280	175	897

Table 27: Planar Façades classified into different categories depending on the number of images that these facades are visible in.

With the before explained workflow, each of the 897 image segments of the 286 facades are processed. Table 28 summarizes the results of the floor and window detection for planar façades. It illustrates the number of images in which the floor counts, and separately the window counts, were correct. In the look angle segment between 15° and 20° one finds 270 facades, and in 232 of those, the floor count was correct. Similarly, in the same look angle range, the window counts were correct in 213 of the 270 image segments corresponding to 86% and 79% detection rate. Since we operate with 2 or more vertical images of each facade, we can exploit the redundancy to find a most likely correct floor and window count.

Angle [deg]	*< 5*	*5 - 10*	*10 - 15*	*15-20*	*20-25*	*> 25*
Floors	0	0	125 / 162	232 / 270	255 / 280	162 / 175
Floor detection [%]	0	0	77	86	91	93
Windows	0	0	109 / 162	213 / 270	232 / 280	151 / 175
Window detection [%]	0	0	67	79	83	86

Table 28: Counting floors (above) and windows (below) from vertical images versus look angles.

These numbers can be condensed into a single figure of merit, for example if one decides on a single floor and window count per building from multiple values in overlapping images. In total, the 286 planar facades form the test data set have 858 floors and 4290 windows.

One finds the following:

Accuracy of building detection	100%	all 216 buildings were detected;
Accuracy of floor count	93%	of all 858 floors were counted;
Accuracy of window count	86%	of all 4290 windows were counted.

The window locations do get defined to within ± 3 pixels in horizontal as well as vertical directions, converting to a value of ± 0.3 m vertically and ± 0.3 m horizontally.

Consideration of the cause of failures finds occlusions as the major culprit, followed by facades deviating from the assumption that they are within a single facade plane, and that windows are unobstructed by balconies or awnings.

The 177 complex facades and the floor and window detection, described in the section before, fails because of the complexity of these facades (see Figure 63). Therefore, in the next part we evaluate the 3D façade reconstruction and the performed floor and window detection based on this reconstruction.

Analysis of the 3D Façade Point Clouds

In Table 29, we have categorized our facades into different classes depending on the number of images used for their reconstruction. As one can see, most of the facades are found in two to four images. Unfortunately, because of the constraints concerning angles there are 7 facades that are visible just in one image. For these facades a 3D reconstruction is not possible. Primarily these are facades that are aligned transverse to the flight direction. We have divided our facades into two categories: complex and planar facades. Complex ones are facades with multiple façade planes (see Figure 60) whereas planar façades have just one major plane (see Figure 63).

	1 image	*2 images*	*3 images*	*4 images*	*5 images*	*total*
No. Façades	7	46	55	34	31	177
No. Images	7	92	165	136	155	555

Table 29: Complex Façades classified into different viewing angle categories

Next we performed a facade plane detection for planar and complex facades using the 2D column-wise approach (see Table 30) and the 3D approach (see Table 31).

	1 image	*2 images*		*3 images*		*4 images*		*5 images*		*total*	
images	17	127		145		104		66		463	
complex	0 %	64%	30/46	74%	41/55	79%	27/34	84%	26/31	70%	124/177
planar	0 %	75%	61/81	77%	69/90	86%	60/70	94%	33/35	78%	223/286
total	0 %	72%	91/127	76%	110/145	84%	87/104	89%	59/66	75%	347/463

Table 30: Evaluation of the facade plane detection for planar and complex facades using the column-wise approach

	1 image	*2 images*		*3 images*		*4 images*		*5 images*		*total*	
images	17	127		145		104		66		463	
complex	0 %	79%	36/46	84%	46/55	79%	27/34	84%	25/31	76%	134/177
planar	0 %	93%	75/81	89%	80/90	97%	68/70	97%	34/35	90%	257/286
total	0 %	87%	111/127	87%	126/145	84%	91/104	89%	59/66	84%	387/463

Table 31: Evaluation of the facade plane detection for planar and complex facades using the 3D approach

In 50% of all cases where façade plane detection fails, the main reason is the combination of images from two different flight strips for the 3D reconstruction. Therefore, one should try to avoid combining images from two flight strips when reconstructing facades. The main reasons are distortions because of the different viewing perspectives that result in noisy matching results. The big problem is that one cannot just use images from one flight strip for façade reconstruction in some cases, because otherwise there are too few images left for 3D reconstruction. In our test area, without combining images from different strips for 20 facades evaluation wouldn't be possible.

When one compares the results of Table 30 with Table 31 one will notice that the detection rate using 3D data is around 10% higher than the 2D detection rate. The main reason for this is that the use of the whole point cloud incorporates much more information about the façade structure for the characterization than the 2D approach. Nonetheless, the results from the 2D approach are better than expected with an overall detection rate of 75%.

Another important aspect is where the facades are located within the vertical aerial images. There are 3 major areas one has to identify: (1) images in close-up range, (2) middle range and (3) far field.

Images in the close-up range are photographed under a small angle-of-view, in our case these are facades visible in 2 images. Middle-range facades are all facades visible in 3 images, and far field facades are facades visible in 4 or more images. Depending on the angle-of-view the quality of the façade details is high or not, which affects the results. As one can see in Table 30 and Table 31, the detection rate for complex facades increases depending on the number of images used.

Next, let's take a look at the geometric quality of the point cloud. For this evaluation we have selected 10 planar and 10 complex facades from every category. Then we choose 2 areas (20 x 20 pixels) in every façade and calculate the standard deviation of the points in this area from the reference plane. The reference plane was defined manually in the aerial images and then projected into the orthophoto for the comparison. Table 32 shows the results of this calculation. As you will notice, the standard deviation decreases as the number of images used for reconstruction increases, which is expected. Using images from different strips the quality decreases by a factor of 5.

	2 images	*3 images*	*4 images*	*5 images*
standard deviation [cm]	19	17	16	13

Table 32: Assessment of the geometric quality of the point cloud

The geometric quality of the point clouds is important when determining the extruding façade masonry, as well as for the enhancement of the building footprints. Table 33shows the results of the floor and window counts based on the 3D approach for all facades, and shows the effect of the number of overlapping images in use.

	2 images		*3 images*		*4 images*		*5 images*		*total*	
	planar	complex	planar	complex	planar	complex	planar	complex	**planar**	**complex**
Floors	83%	86%	89%	83%	90%	84%	93%	90%	**88,9%**	**85,8%**
Windows	68%	75%	79%	80%	86%	81%	88%	82%	**82,9%**	**80,3%**

Table 33: Evaluation of Floor and Window Detection (planar and complex facades)

The accuracy of counting the number of floors and windows increases with the number of available images per façade. One reason for this is the help of overlaps in resolving ambiguities that occur because of occlusions and distortions. Another reason is the change of the angles-of-view.

One can see in Table 33 the accuracy for floor and window detection increases as the number of available images per facade increases. One reason for this is that redundant image data can improve the results by solving ambiguities that occur because of occlusions and distortions. Another reason for the increase in the detection rate is the change of the angle-of-view.

For floor detection we achieve a success rate for planar facades of 88,9% and for complex facades of 85,8%. For window detection the success rate for planar facades lies around 82.9% respectively 80.3%. Reasons for misclassifications are partial shadows, occlusions from vegetation and other buildings, distortions due to the steep look angle and difficult façade structures. Without the use of the 3rd dimension almost none of these complex facades could be characterized correctly.

Concerning the segmentation of complex facades into different classes like balconies, bay windows and other emerging façade structures we achieve high accuracies. The 2D approach achieves a detection rate of 66% for balconies and 0% for bay windows. The reason why none of the bay windows could be detected is because the column-wise plane detection neglects small emerging structures. The 3D approach achieves a detection rate of 82% for balconies and 72% for bay windows. Reasons for misclassifications are distortions in the point clouds and occlusions.

Oblique versus Vertical Imagery

In the prior section, we have already performed a comparison between vertical and oblique aerial images based on radiometry, geometry and visual impact. In this section, we want to evaluate both techniques in connection with floor and window detection. The experiments in this section were performed using the Graz test data set that was acquired using an UltraCamX with an 80/60 overlap and a GSD of 10cm. The oblique images were taken from the Microsoft BING/Maps website in its "Classic" version, and have a GSD of nominally 12cm.

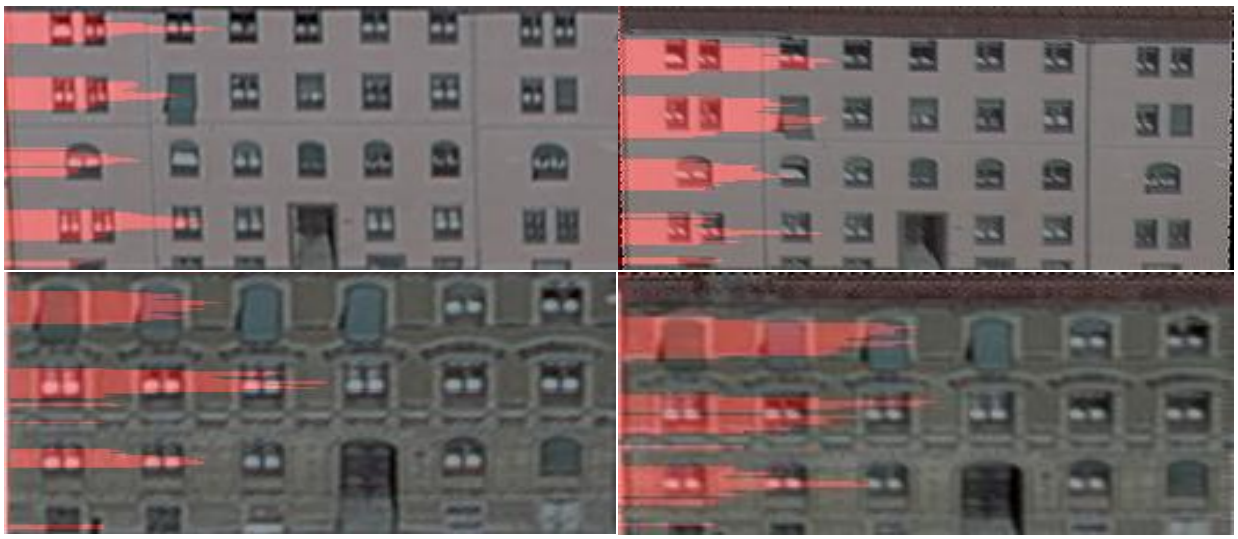


Figure 74: Two examples for floor detection using edge histograms. (left) oblique aerial image (right) vertical aerial image;

Counting Floors

A less subjective and more quantitative comparison of oblique versus vertical is expected to result from analysing the images and extracting semantic information.Figure 74 explains that a histogram is being built from horizontal Prewitt edges and local extrema of the histogram serve to get a floor count. Applying this approach to about 870 facades in the Graz test area's vertical images and to a subset of 120 facades in the corresponding oblique images (from Bing/Maps) leads to table 34. We find in this type of quantitative analysis that the result is seriously compromised by the occlusions which naturally are larger in the oblique images.

Counting Windows

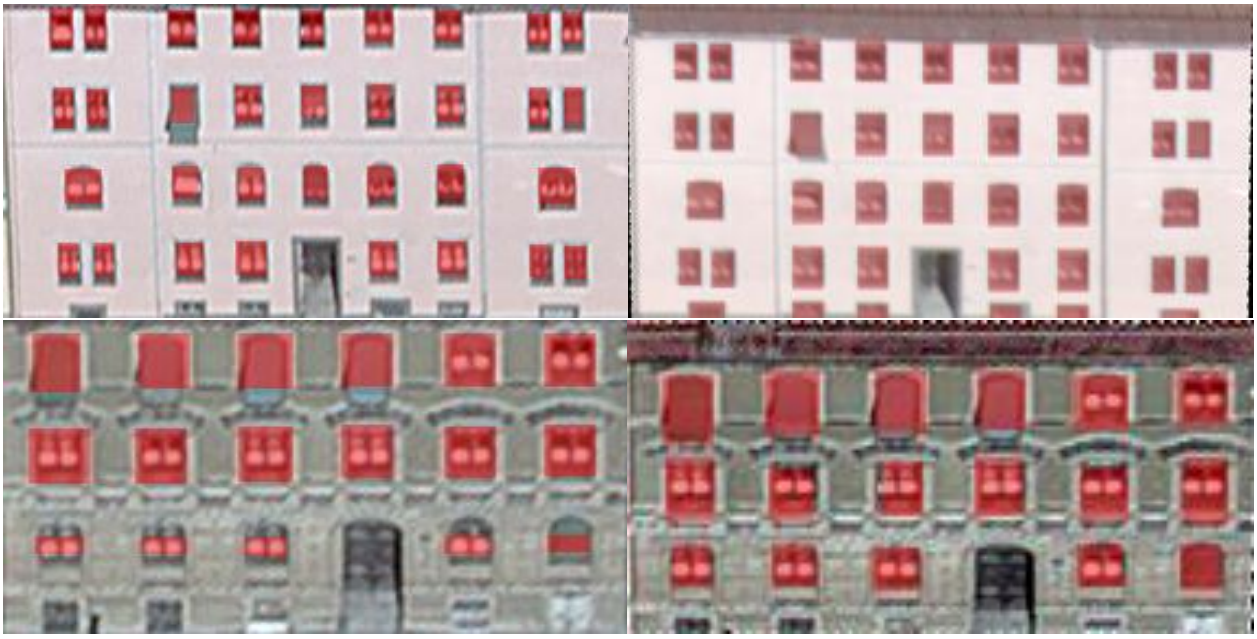
A histogram-based count can also deliver the number and locations of windows. Figure 75 explains the principle of the approach. Of course one will want to apply various constraints on window size and distance between etc. to overcome the effects of data noise.

Table 34 shows the accuracy achieved in the Graz test data set for the facades on vertical and oblique images. Again, occlusions are the main obstacle to a competitive result from oblique images.

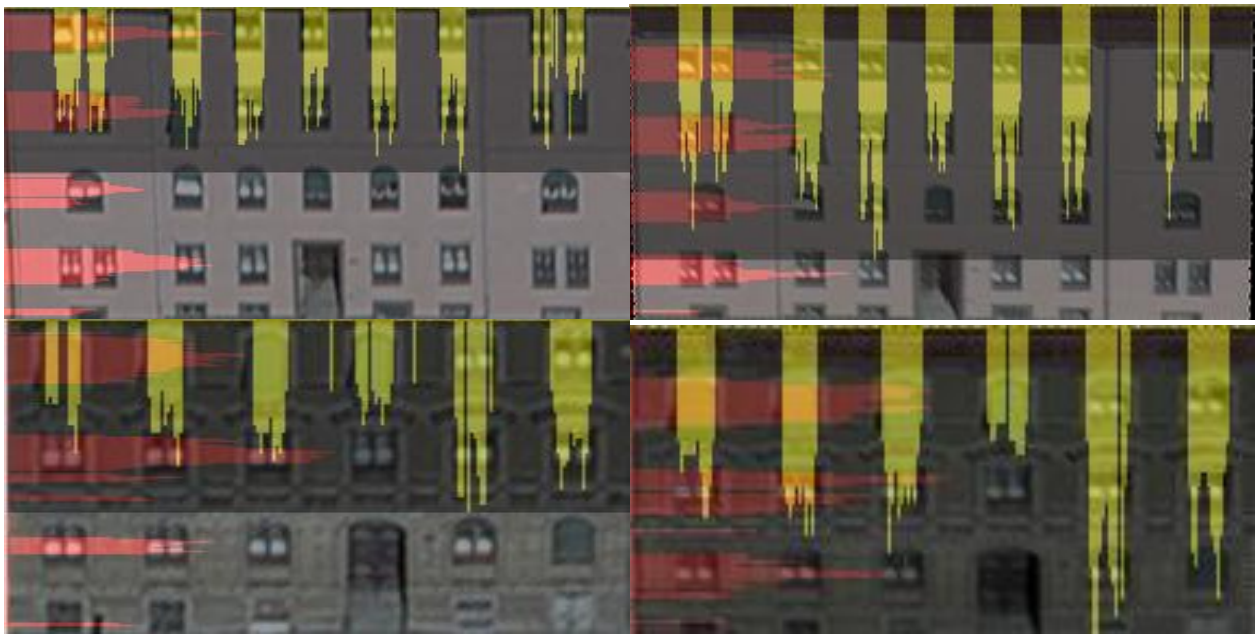
Angle [deg]	*< 5*	*5 - 10*	*10 - 15*	*15-20*	*20-25*	*> 25*	*Oblique*
Floors	0	0	125 / 162	232 / 270	255 / 280	162 / 175	90/120
Floor detection [%]	0	0	77	86	91	93	75%
Windows	0	0	109 / 162	213 / 270	232 / 280	151 / 175	79/120
Window detection [%]	0	0	67	79	83	86	66%

Table 34: Counting floors (above) and windows (below) from vertical images and results depending on look angles. Last column is from oblique images where floor counts are compromised by occlusions.

In Table 34 the success rate of window and floor detection is calculated by dividing the total number of facades for every angle (e.g. 10-15) by the number of facades where the floors and windows are correctly determined (e.g. 109/169). As one can see, the floor and window detection results for oblique images are not as good as the results using vertical aerial images. Reasons for that are the poor resolution of the oblique aerial images and occlusions from other buildings and vegetation. The main reason for the results in connection with the floor detection are occlusions. One of the reasons for the outcome of the window detection is the poor resolution of the images.



(a) detected windows (oblique aerial image) (b) detected windows (vertical aerial image



(c) horizontal and vertical edges (oblique aerial image) (d) horizontal and vertical edges (vertical aerial image)

Figure 75: Window detection approach using edge histograms. Images a and b illustrate detected windows using oblique and vertical aerial images; Images (c) and (d) show vertical and horizontal summarized and normalized edges

Discussion

We demonstrate that façades can be analyzed with a 90% success rate from vertical aerial photography. This is feasible since the images have been taken with large overlaps so as to image each façade at a sufficiently large look angle of 20° to 27°. We also show that the visual inspection of vertical versus oblique images favors the vertical data due to better radiometry at comparable pixel sizes. The major problems of oblique images are occlusions that prevent one from counting the correct number of floor and windows.

The efficiency of aerial imaging may favor vertical technologies over the oblique approach. Vertical images today produce 280 Megapixels per exposure, whereas oblique cameras still operate at the 10 Megapixel level. Even if one were to consider that in a Maltese Cross arrangement one operates with 5 such cameras, this still adds up to only 50 Megapixels. A limitation of the current "normal angle" aerial cameras is the look angles one can achieve in the direction of flight at perhaps 17° off nadir. Solutions are either a cross flight pattern, or the use of a wide angle camera model such as the UltraCam Xp-WA with 26° in flight direction, or the use of the new single CCD-chip DMC-II, recently announced by Intergraph.

Going beyond a mere "eye candy" approach for the use of oblique images, one will quickly find that novel high-redundancy vertical aerial images offer a superior source of information about urban areas, street canyons and facades. We suggest that benefits from vertical aerial photography have been undervalued, and that benefits from oblique images have been overstated.

7.6. Conclusion

While one will expect that 2D measurements of surface areas can be obtained from vertical aerial photography, one might be uncertain about the value of such vertical imagery when analyzing façades. However, our work shows that vertical aerial photography is a valuable source of façade-related information, both in terms of planar facades as well as in the event that facades are complex and require a 3D approach.

Initial work focused on planar facades, and resulted in a correct floor count at a rate of 93% and a correct window count at 86%. A study of the dependence of these counts, on the angle off-nadir in the images, reveals that best results are achieved in the range of 20° to 27° off-nadir. At greater angles, such as those used in oblique aerial photography, occlusions start to become distinctive so that floor and window counts fail. Such counts will also fail if a façade is complex, yet the analysis method is unaware of that complexity. One will need to consider the 3D shape of building facades, for example with extruding stair cases, balconies, awnings. We explain in this paper an approach based on 3D clouds of façade surface points, all extracted from overlapping vertical images.

We show that the use of the 3rd dimension when characterizing building facades strongly improves the results. We achieve success rates of 85.8% for floor detection and 80.3% that were unfeasible without reconstructing a façade in 3D. We also show that it is possible to segment a

complex façade into different classes like balconies, bay windows and other emerging façade structures.

There are several avenues for improvements of the detail extraction from complex facades, and improvements of our understanding of how well this works. First, are data experiments in the form of a study with vertical aerial images with different GSD and different overlaps. Second, is the ability of automatically recognizing occlusions and then responding to occluded facades by taking advantage of the overlapping images and their multitude of look angles. Innovations in window recognition will then become relevant, as will site-dependent approaches, to architectural styles.

CHAPTER EIGHT

ROOF DETAIL CHARACTERIZATION

Contents

8.1	Introduction	171
8.2	DSM Smoothing	172
8.3	Plane Detection	173
8.4	Classification	177
8.5	Skylight Detection	185
8.6	Roof Characterization	186
8.7	Experiments	192
8.8	Discussion	211

8.1. Introduction

Building roofs play a major role for the generation of 3D city models or the solar cadastre. Both the roof shape and the existence of superstructures on a roof are of interest. We want to answer the following questions: What is the roof shape of a building? Is the rooftop extended? How accurate? Type of building? In this chapter, we want to discuss the necessary processing steps to make a characterization of a roof feasible.

Roofs and their superstructures need to get mapped using the building's point cloud and image segments. The process consists of five steps (Figure 76). First, is the smoothing of the noisy digital surface model (DSM) using total generalized variation TGV presented in [Pock et al., 2008]. Second, is the extraction of all plane roof segments. In this work we introduce two approaches for plane detection. The first approach called J-Linkage was introduced in [Toldo et al., 2008] and resembles the RANSAC algorithm for plane detection, and includes the consideration of neighborhoods. The second approach uses a multi-label image segmentation minimizing labeling costs for plane detection. In the latter case a DSM smoothing is optional. After the plane detection, each plane surface patch gets labeled as a major, minor or superstructure plane. This initially uses the DSM point cloud and then refines the segmentation by incorporating the RGB-

image data. In a region growing approach, plane segments get merged into larger surfaces or are set aside for further analysis.

A fourth process element is the search for skylights in the major and minor roof planes using a template matching method in the form of fast directional chamfer matching introduced by [Liu et al., 2010].

The fifth step is the characterization of the data, and assignment of labels to the isolated data elements and collected list of features. The roof gets an overall attribution to a specific architectural style; the "left-over" data are the basis for a classification into dormers, chimneys and sky lights.

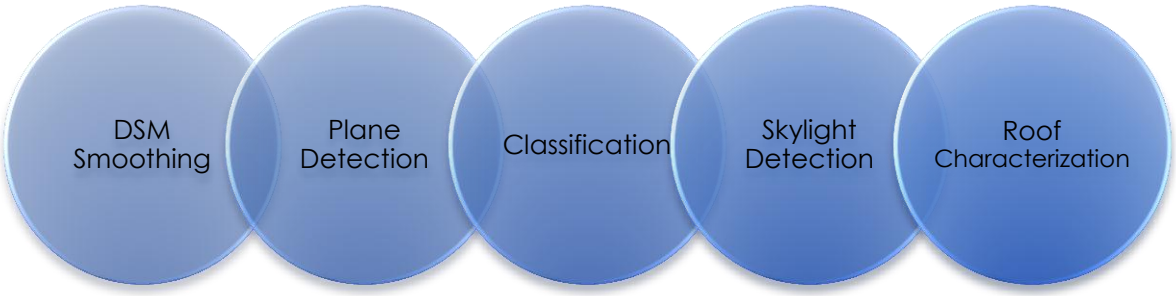


Figure 76: Sequence of the five major processing steps

8.2. DSM Smoothing

Photogrammetrically measured elevation data (range data) are noisy. At a pixel to multi-pixel level, this will obstruct any search for plane surface patches. For this reason, the 3D point clouds need smoothing. The filtering of elevation measurements is a traditional function of photogrammetric work in the form of an approximation and interpolation of terrain elevation grids from manually measured surface points. In the current application, we have very dense data and need a filter that preserves sharp surface discontinuities and small structures while producing smoothness within the roof. A recent innovation is "total generalized variation" TGV introduced by [Pock et al., 2008].

The basic idea of the variational approach is that the solution of the model is given by the minimizer of an energy-functional that is usually composed of two terms. The regularization term copes with the a-priori assumption about the smoothness properties of the solution, and the data term forces the solution to be similar to the input data. The second order variant of this functional is ideally suited for piecewise affine surfaces and therefore an ideal case for buildings that can be approximated by piece-wise planar surfaces.

The data term is based on the robust Huber-L^1 norm; the regularization term is based on TGV. The strength of the Huber-L^1 norm is that it reduces the staircasing effect that other models create, and it better reflects the noise model of real range images. The Huber model is obtained by implementing the Huber norm to the data and regularization term

$$\min_{u} \left\{ \alpha \int_{\Omega} |\nabla u|_{\varepsilon} \, dx + \sum_{l=1}^{K} \int_{\Omega} |u - f_l|_{\delta} \, dx \right\}$$

The main property of TGV is that it allows one to reconstruct piecewise polynomial functions of arbitrary order. It has the nice property to be convex, which means that a global optimal solution can be determined. For the approximation of buildings, it is sufficient to use the TGV regularization of the second order since they can be approximated by piecewise planar surfaces. This method combines TGV regularization of second order with the Huber-L^1 norm

$$\min_{u,v} \left\{ \alpha_1 \int_{\Omega} |\nabla u - v| \, dx + \alpha_0 \int_{\Omega} |\varepsilon(\mathrm{v})| \, dx + \sum_{l=1}^{K} \int_{\Omega} |u - f_l|_{\delta} \, dx \right\}$$

The numerical algorithm is then derived based on a primal dual formulation. For a more detailed description please refer to [Pock et al., 2008]. Figure 77 shows the result of this method for a building of a test dataset from the city of Graz, Austria.

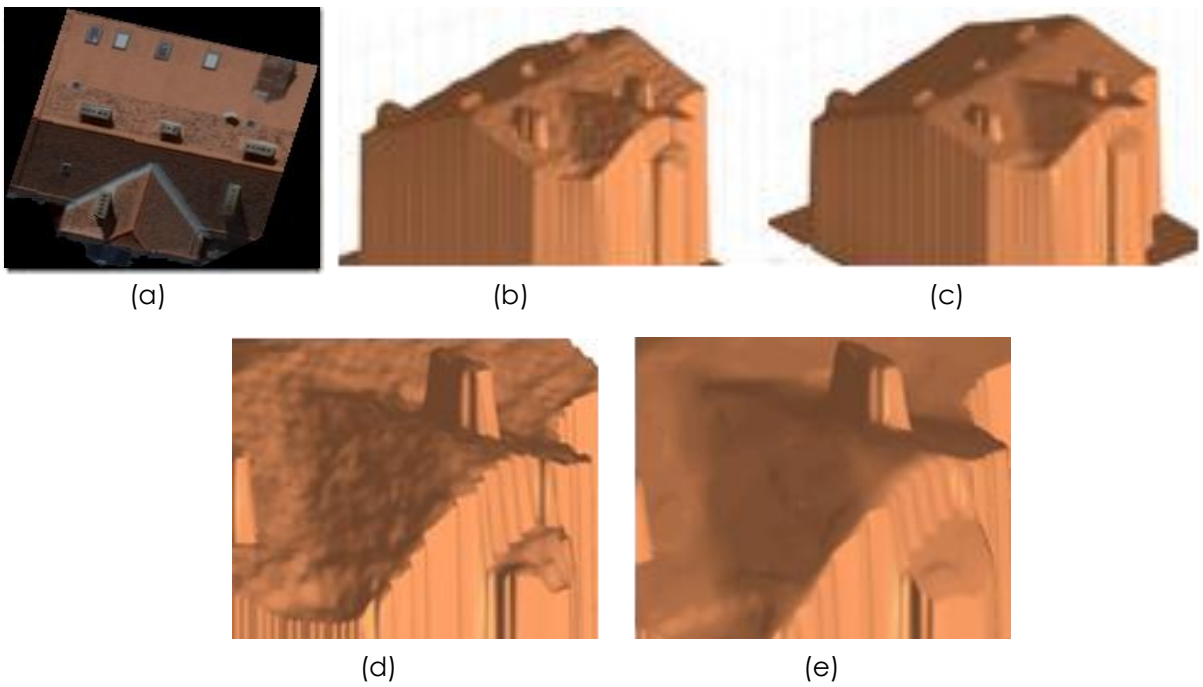


Figure 77: (a) RGB image of single building of the Graz test dataset; (b) Raw range image prior to smoothing; (c) Range image using TGV; (d) Detail of raw range image; (e) Smoothed detail.

8.3. Plane Detection

Roof planes get detected in the smoothed point cloud. Of the many different methods available for the task, RANSAC plane detection has become a standard, in particular in the LiDAR literature [Förstner et al., 2010; Comaniciu et al., 2002; Brenner et al., 2001]. Its principle is well explained in [Fischler et al., 1981; McGlone et al., 2004]. A limitation of RANSAC results

from its disregard for neighborhoods; it treats all points the same without considering that a point might be near another. Moreover, it works only for single structures, which is a major drawback. The problem is that the robust estimator must tolerate both gross outliers and pseudo outliers. To avoid this problem, a sequential RANSAC and multi-RANSAC algorithm was introduced by [Zuliani et al., 2005]. The drawback of these methods is that the number of models that the algorithm detects is user specified and that is not acceptable for our specific application. Another common method is Randomized Hough Transform (RHT) that is based on random sampling and voting [Xu et al., 1993]. A histogram is generated over the parameter space, and models are randomly determined and recorded in the histogram. Detected peaks in the histogram correspond to the sought model. Its major advantage over the other methods is that it can easily be adopted for multiple structures using multiple peaks in the parameter space. Compared to RANSAC, RHT doesn't need the number of models in advance. However, the method has typical disadvantages of Hough Transform methods such as limited accuracy and low efficiency. Moreover, the choice and discretization of the parameter space is crucial for the outcome of the algorithm.

Another alternative is mean shift [Comaniciu et al., 2002] where data is mapped into the parameter space through random sampling and then seek the modes of the distribution. Like RHT, the parameterization is crucial and not an intrinsically robust technique.

A recently proposed method [Zhang et al., 2006] is based on the analysis of the distribution of residuals of individual data points. The necessary hypotheses are generated by a RANSAC-like sampling process. One of the major disadvantages of this method is that the peak corresponding to a given model becomes less localized as the point-model distance increases.

For this project, we have evaluated three different plane detection methods:

- Classical RANSAC APPROACH
- RANSAC like approach that incorporates neighborhood relations (J-Linkage)
- Multi Label Image Segmentation

Classical RANSAC Approach

Random Sample Consensus (RANSAC) is an effective method to robustly fit models to noisy data. The model hypotheses are computed by using a minimum number of data points and the score of all matches is calculated.

Algorithm: Classical RANSAC Approach; Workflow from [Fischler et al., 1981; Irschara, 2012]:

- Randomly select a minimal subset of n data points from the whole set N of observations and compute the model from the subset only.
- Determine the set of data points $N_i \subseteq N$ which are within a distance threshold t of the model. The consensus set N_i defines the inliers of N.
- If N_i contains more inliers than some threshold T, re-estimate the model using all the points in N_i and terminate.

- If the size of N_i is less than T, select a new subset and repeat steps 1,2,3.
- After a maximal number of N trials with $N = f(|N_i|)$, the largest consensus set N_i is selected and the model is re-estimated using all the points in the subset N_i.

J-Linkage Approach

To overcome all drawbacks mentioned before, [Toldo et al., 2008] introduced a method that does not work in the parameter space like RHT, nor in the residual space like [Zhang et al., 2006] but rather in the conceptual space. It is a variation of RANSAC that incorporates neighborhood relations.

It starts by random sampling to generate model hypotheses. Minimal sets are constructed in a way that neighboring points are selected with higher probability. This means that if a point x_i has already been selected, that x_j has the following probability of being drawn:

$$P(x_j|x_i) = \begin{cases} \frac{1}{Z} exp - \frac{\|x_j - x_i\|}{\sigma^2} & if\ x_j \neq x_i \\ 0 & if\ x_j \neq x_i \end{cases}$$

where Z is the normalization constant, σ is a heuristically chosen constant, x_i and x_j are single observations. After the creation of all hypotheses, a preference set of preferred hypotheses is created for each point. Points belonging to the same structure have a similar preference set, thus they are close in the conceptual space. Each point is represented with the characteristic function of the set of models preferred by that point. Multiple models are shown as clusters in the conceptual space.

To find the models, the method uses an agglomerative clustering procedure. This method works in a bottom-up manner where two clusters with the smallest distance are merged. The result of this calculation varies depending on the used method to calculate the distance. J-Linkage uses the Jaccard Distance to compute the distance between the respective preference sets X and Y.

$$d(X,Y) = \frac{|X \cup Y| - |X \cap Y|}{|X \cup Y|}$$

It measures the degree of overlap of the two data sets and ranges from 0 (identical sets) to 1 (disjoint sets). Because we are just interested in elements whose preference sets overlap, we set the cut-off value to 1. Using this feature, clusters of points have the following properties:

- There exists at least one model that fits all the points of the cluster
- Two distinct clusters cannot be fitted by one model

After all models are generated in a last step, they calculate the final model of each cluster of points using least square fitting.

Figure 78 shows the result of the plane detection for one building of our test dataset. The comparison of a plane with the point cloud produces a measure of accuracy. We find that the test data have elevation noise of ±12cm; this is to be related to the GSD at 10 cm.

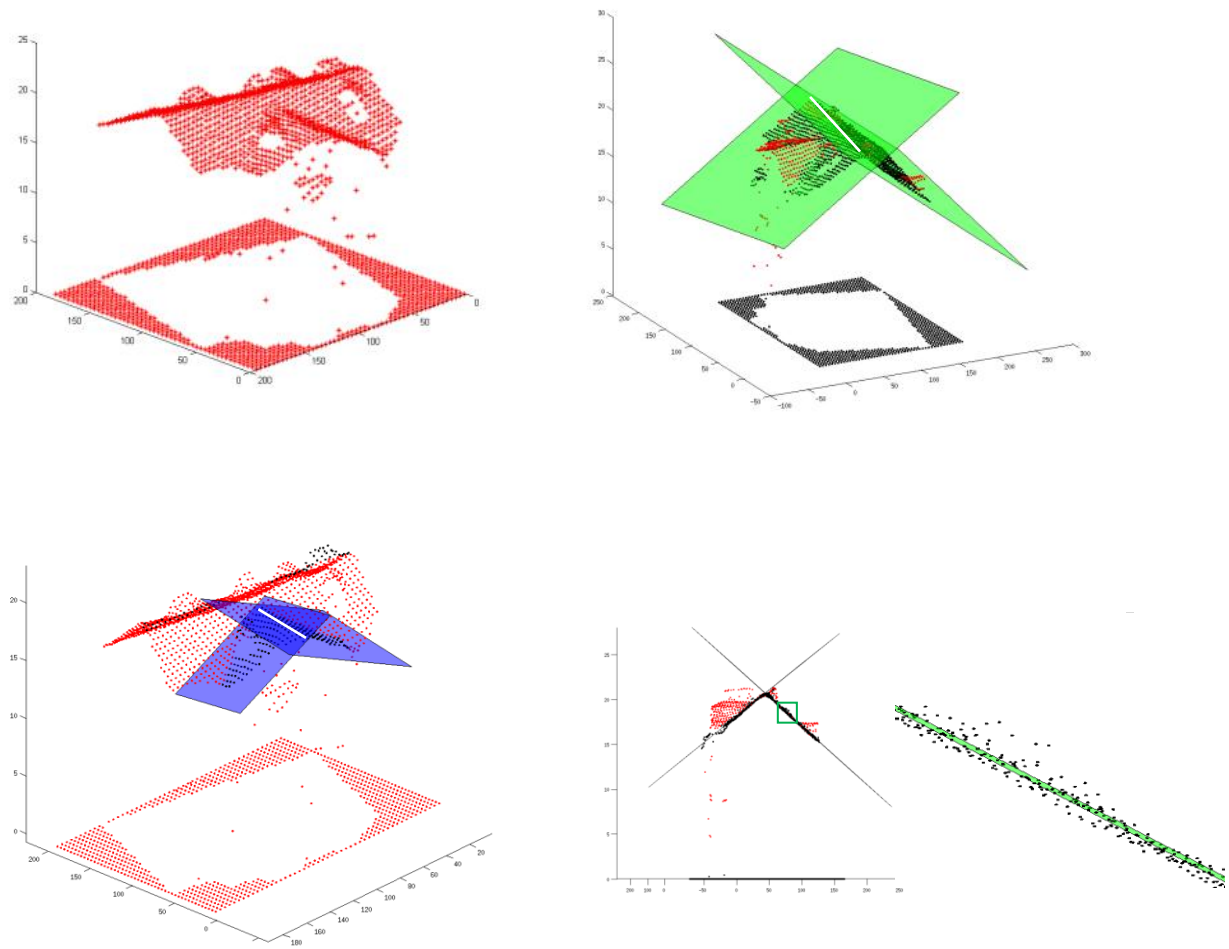


Figure 78: Axonometric view of (a) point cloud of building roof, points in red; (b) point cloud and detected major planes in green. Note the intersection line and points belonging to the planes in black (c) point cloud and detected minor planes in blue. Note intersection line and points belonging to the planes in black; (d) side view of point cloud and major planes; at right is an enlarged detail showing standard deviation between points and plane at ±12cm.

Multi Label Segmentation

We focus on the general minimal partition problem for an arbitrary number of problems.

$$\min_{\Omega_i}\left\{\frac{1}{2}\sum_{i=1}^{K} Per(\Omega_i) + \sum_{i=1}^{K}\int_{\Omega_i} f_i(x)\,dx\right\}$$

$$s.t.\ \Omega = \bigcup_{i=1}^{k}\Omega_i,\ \ \Omega_i \cap \Omega_j = 0\ \ \forall_i \neq j.$$

Here f_i are unary potentials that are integrated inside the region Ω_i. The regularization term penalizes the contour length, and is weighted with ½ to account for double counting.

Label Costs

This subchapter was reprinted from [Unger 2012]. *We extend the minimum problem of (fast relaxation)*

$$\min_{u_i}\left\{\sum_{i=1}^{K}\int_{\Omega}|\nabla_{u_i}|+\sum_{i=1}^{K}\int_{\Omega_i}u_i(x)f_i(x)\,dx\right\}$$

with an additional label cost term, as proposed by [Yuan et al., 2010].

$$\min_{\Omega_i}\left\{\sum_{i=1}^{K}Per(\Omega_i)+\sum_{i=1}^{K}\int_{\Omega_i}f_i(x)\,dx+\gamma\|\boldsymbol{I}_{\Omega_i}\|_{\infty}\right\},$$

$$s.t.\ \Omega=\bigcup_{i=1}^{k}\Omega_i,\ \ \Omega_i\cap\Omega_j=0\ \ \forall_i\neq j.$$

The infinity norm penalizes the maximum value of the characteristic function $\boldsymbol{I}_{\Omega_i}$, *and minimizes the number of non-empty segments. We again use the characteristic functions* $u_i:\Omega\rightarrow\mathbb{R}$ *to represent the regions* Ω_i.*The label cost term then becomes* $\sum_{i=1}^{K}\gamma\|u_i\|_{\infty}$.

In order to optimize the label cost term, we follow the ideas proposed in [Yuan et al., 2010]. To replace the L^{∞}*-norm in* $\|u_i\|_{\infty}$, *the scalar t is introduced. We then arrive at the equivalent representation*

$$t,\qquad s.t.\,u_{k,l}\leq t,\forall k,l.$$

The pixel wise constraints can be written in a single term when using a primal-dual formulation $\langle q,P\binom{u}{t}\rangle$. *Where the Lagrange multiplier* $q\in\mathbb{R}^{MN}$ *has to be positive to account for the inequalitiy constraint. We define the* $(MN+1)\times MN\ matrix\ P=(I,-\boldsymbol{I}^T)$, *with MN x MN identity matrix I and the row vector* $\boldsymbol{I}$ *of size MN with all entries equal to 1.* For a more detailed description of this method please refer to [Unger 2012].

8.4. Classification

The roof plane detection is the basis for a classification of these detected planes into different categories depending on different parameters. We divide them into three classes: major planes, minor planes and superstructure planes. We further divide the category superstructures into chimneys, dormer windows and other structures. The required classification workflow, illustrated in Figure 79, consists of three major processing steps:

- Initial Classification
- Superstructure Classification
- Refined Classification

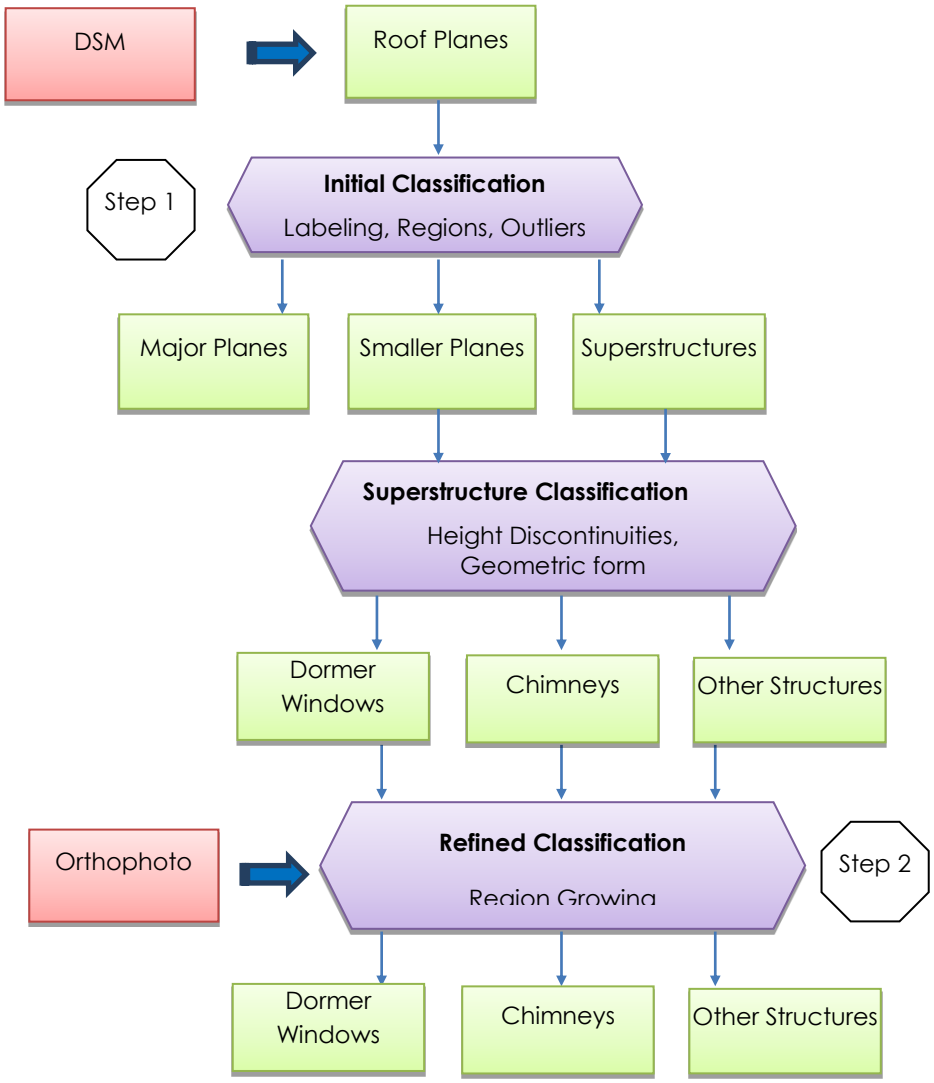


Figure 79: Proposed classification processing steps divided into two major steps.

Initial Classification

The initial classification analyzes every plane segment to determine if it is linked to any other segment to build a larger "region". Then the size of the resulting region or assembly of plane segments gets considered. Depending on the size of the regions, with respect to the overall size of the roof, we assign each region to an appropriate category. Additionally, the location of the respective plane on the roof is also taken into consideration for the classification.

A differentiation between major and minor or smaller planes is heuristic and a function of architectural customs in an area. The threshold can be assigned as a function of every building's footprint and the size of the individual plane segments. But some smaller planes will get erroneously assigned to the major plane class and vice versa. Figure 80 illustrates the result of the initial classification for four example buildings.

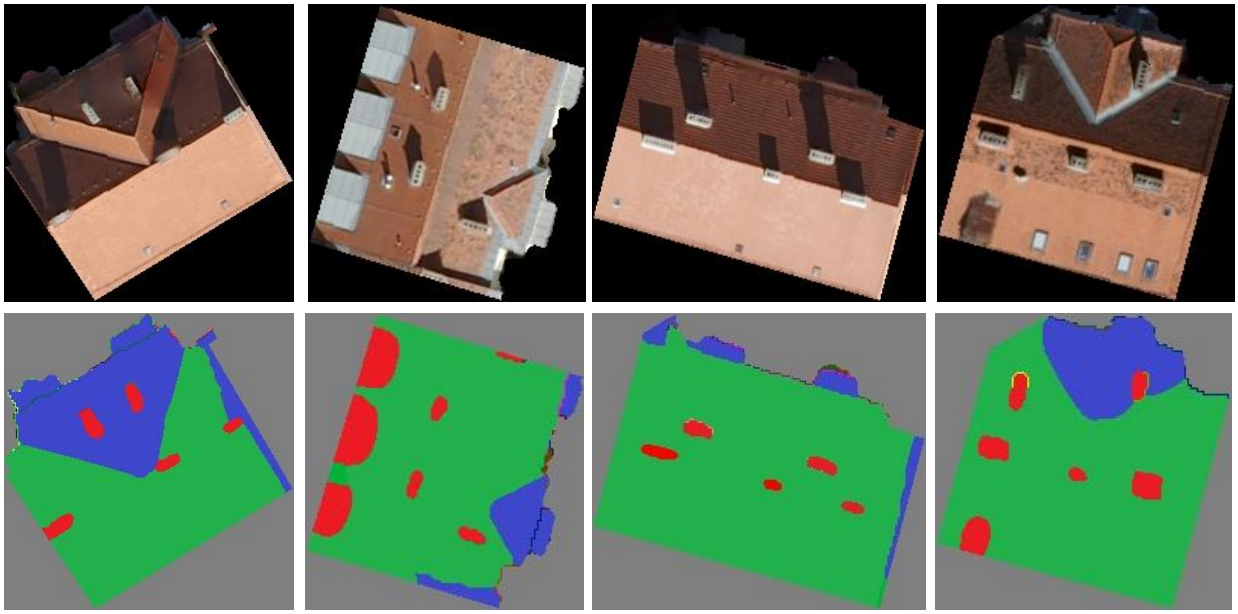


Figure 80: Initial classification result for four buildings (green: major roof planes; blue: smaller roof planes; red: superstructures)

Superstructure Classification

A secondary analysis addresses the smaller planes and superstructures. Depending on their height values, with respect to the neighboring pixels, these regions are classified as part of a smaller plane segment or of a superstructure or are eliminated entirely. To be able to distinguish smaller planes and superstructures we have to know how superstructures are defined. We use a definition from Wikipedia that defines superstructures as follows[28]:

"*A **superstructure** is an upward extension of an existing structure above a baseline. This term is applied to various kinds of physical structures such as buildings, bridges, or ships.*"

This means that superstructures are structures with height discontinuities at their edges that protrude from a planar surface. To achieve meaningful results, we distinguish between height discontinuities at the edges of the roof and within the roof by using the information about a building from the building classification. With this additional information, it is possible to differentiate between smaller planes and superstructures.

In a next step, we want to classify the superstructures into three different types: chimneys, dormer windows and other structures. In a first step, we want to find definitions for chimneys and dormer windows. We use two definitions to describe a chimney. The first definition comes from Wikipedia and defines a chimney as follows[29]:

"*A **chimney** is a structure for venting hot flue gases or smoke from a boiler, stove, furnace or fireplace to the outside atmosphere. Chimneys are typically vertical, or as near as possible to*

[28]http://en.wikipedia.org/wiki/Superstructure
[29]http://en.wikipedia.org/wiki/Chimney

vertical, to ensure that the gases flow smoothly, drawing air into the combustion in what is known as the stack, or chimney, effect. "

The second definition coming from thefreedictionary.com[30] and adds more crucial information about a chimney:

"*A passage through … The part of such structure that rises above a roof.*"

For chimneys one postulates that on all 4 edges of the region are height discontinuities. The specifications for chimneys differ for every country. In Austria and Germany, for instance, the following requirements have to be fulfilled.

The openings of an exhaust system have to surmount the ridge by at least 40cm or have to be at least 1m away from the next roof plane (see Figure 81). Depending on the kind of heating system, it is also feasible that the opening of a chimney is up to 40cm away from the next roof plane (see Figure 81). Chimneys usually have smaller and narrower forms with a maximum width of 50cm and can be quadratic or elongated depending on the number of openings.

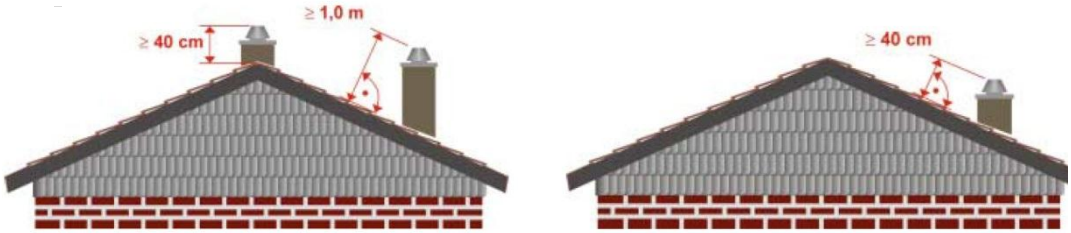


Figure 81: Requirements for chimneys on building roofs; taken from[31]

To define dormers we use a definition from Wikipedia that describes this superstructure in the following way:[32]

"*A **dormer** is a structural element of a building that protrudes from the plane of a sloping roof surface.*"

Dormers have height discontinuities on at least three edges and are in general larger than chimneys. Figure 82shows five possible types of dormer windows. Within this project we classify dormers into these five categories. However, we know that there exist a huge number of different dormer types depending on the region.

[30]http://www.thefreedictionary.com/chimney
[31]http://www.ralf-kampa.de/bilder_rlp/files/din18160.pdf
[32]http://en.wikipedia.org/wiki/Dormer

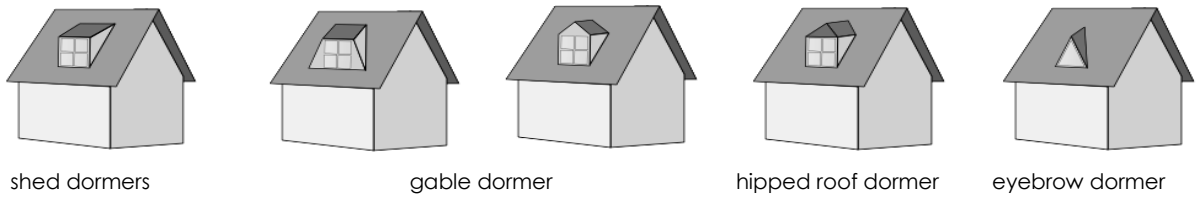


Figure 82: Different types of dormer windows [taken from Wikipedia]

All structures that don't fit into these two categories like satellite antennas, walking paths on a roof etc. are classified as other structures. Figure 83 illustrates the results for the superstructure classification step.

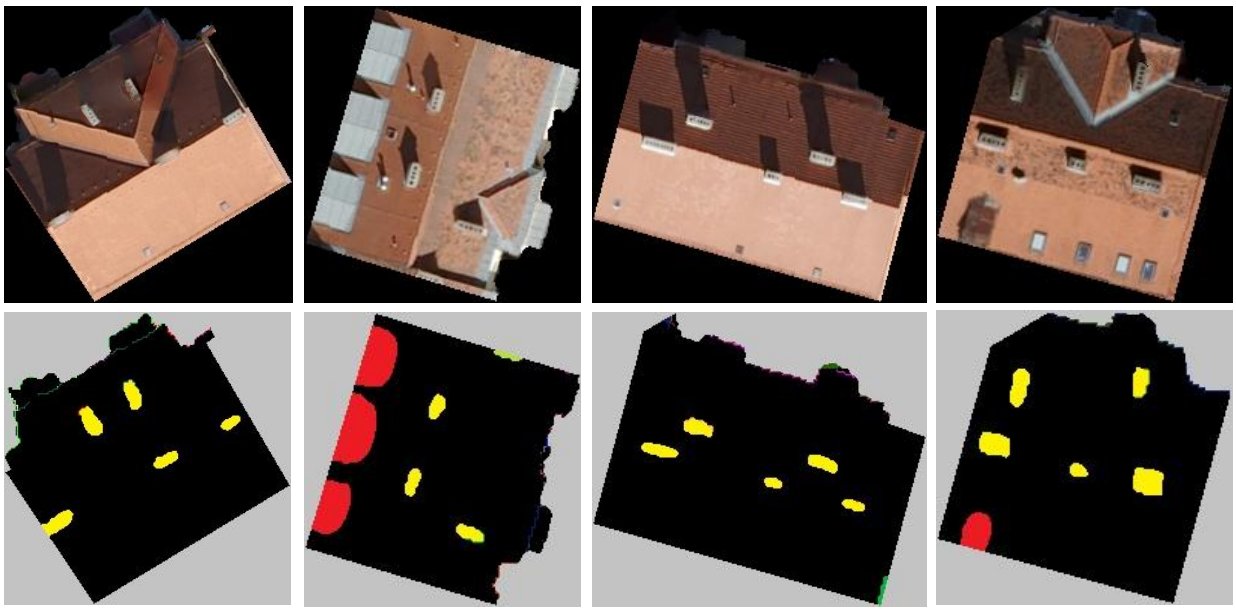


Figure 83: Superstructure classification for four buildings based on initial classification from Figure 80. yellow: chimneys; red: dormer windows; green: other structures

Refined Classification

The classification result based on the plane detection shows rough plane edges. However, for superstructures we are interested in accurate borders that delineate the structures from the other planes. Chimneys, for instance, have rectangle or quadratic shapes instead of round shapes like in Figure 83. We perform a region growing based on elevation and RGB-values for a more accurate and robust segmentation of these roof parts as proposed by [Koivunen et al., 1991].

Region Growing

Region growing is a region based image segmentation with its main goal to split an image into different regions according to similarity constraints. Region growing segmentation has the following five characteristics:

- Every pixel belongs to one region

$$\bigcup_{k=1}^{n} R_k = R \; for \; all \; k = [1,2, \dots n]$$

- All points within a certain region must be connected in some sense

R_k is a connected region

- All regions must be disjoint

$$R_k \bigcap R_j = [\quad]$$

- Properties must be satisfied by the pixels in a segmented region

$$p(R_k) = TRUE$$

- Two adjacent regions R_k and R_j must have different properties

$$p\left(R_k \bigcap R_j\right) = FALSE$$

The method starts by choosing an arbitrary seed pixel and comparing it with all neighboring pixels. The seed pixel locations are the centroid of each segment that were determined during the initial classification step. The region is iteratively grown by comparing all unallocated neighboring pixels to the region based on two criteria:

- Similar pixel intensities coming from true orthophoto
- Similar elevation values coming from DSM

The difference between a pixel's intensity value and the region's mean is used as a measure of similarity. Therefore the region mean$\bar{x}$ and the standard deviation σ of the NxN neighborhood is calculated.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{x} = +\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

where x_i is the intensity value of i-th pixel in the NxN neighborhood and n is the total of pixels. The pixel with the smallest difference measured this way is allocated to the region. The pixel is included if it satisfies one of the following two conditions [Romberg et al., 1997]: If the gradient of the pixel is less than 95% of the equalized histogram and the intensity values of the pixel is less or equal to a heuristically defined threshold. If the gradient of the pixel is more than or equal to 95% of the equalized histogram and the intensity levels or the pixel is not more than or equal to one standard deviation away from the region mean. In our case, the seed pixel grows towards its eight surrounding neighbors (see Figure 84).

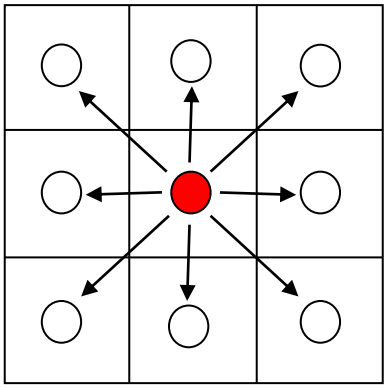


Figure 84: The seed pixel and its 8 surrounding neighbors that are used for region growing

This process stops when the one pixel doesn't fulfill these requirements. When the growth of one region stops, another seed pixel is chosen that doesn't belong to any region and the process starts again until all pixels belong to one region.

The advantage of region growing is that it often gives good segmentations that correspond well to the observed edges. The result of the region growing delivers a refined classification of the superstructures. However, especially for chimneys, the resulting edges are still very scraggy and don´t correspond to reality where they have a rectangle shape. To improve these segmentation results, in a last step, a rectangle fitting is performed. The question is: What rectangle is the best fit for a given set of points?

Rectangle Fitting

Given a set of points the task is it to determine the rectangle that fits the best to all of them. Rectangle fitting is a widely studied problem in computer vision [Chaudhuri et al., 2007; Abellanas et al., 2004; Alt et al., 2001; Suesse et al., 2001]. We apply a method introduced by [Chaudhuri et al., 2007], which fits a bounding rectangle to closed regions. They introduced a five step approach to solve this problem (see Figure 85). In a first step, they compute the boundary of each object and compute its centroid. After that the major and minor axes are determined as well as the upper and lower furthest points with respect to both major and minor axes. By using these points, in a last step, the four vertices of the bounding rectangle can be computed. For a more detailed description about the used algorithm refer to [Chaudhuri et al., 2007].

In a last step, the quality is determined by calculating the percentage of labeled points within the rectangle. If it is larger than a certain threshold (<90% of points are inliers) a different method for rectangle fitting is applied that optimizes the fitting rectangle, this method was introduced by [Abellanas et al., 2004].

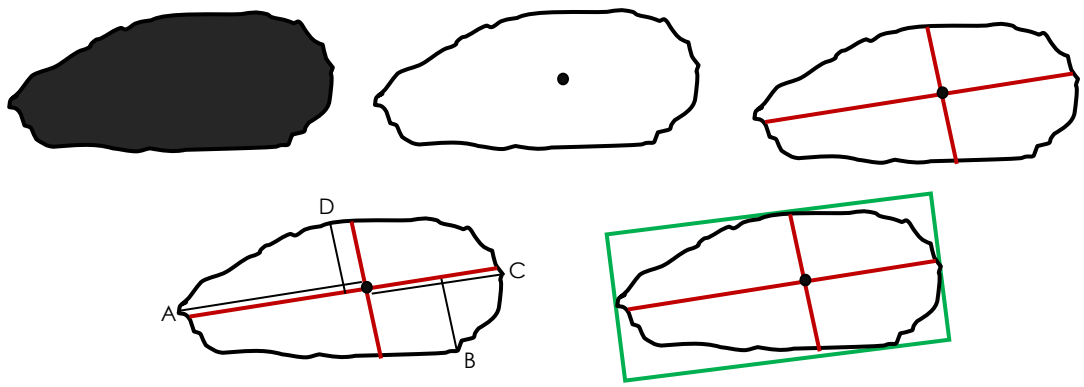


Figure 85: Contour of one extracted chimney; Illustrated are different steps of bounding fitting rectangle: (a) shape of chimney after region growing; (b) centroid and board of the chimney; (c) major and minor axes of area; (d) upper and lower furthest points; and (e) four vertices of bounding rectangle.

In this context, rectangle means the boundary of the rectangle and not the interior. The distance between a point x_i and a rectangle R (boundary) is

$$d(x,R) = \min_{q \in R} d(x,q)$$

Where d(p,q) denotes the distance in the underlying metric. For a subset, S, of admitted rectangles the best fitting one is looked for such that $\max_{x \in R} d(x,R)$ is minimized all over all rectangles R and the best fitting rectangle R_{best} fulfills the following condition

$$\max_{x \in P} d(x, R_{best}) = \min_{R \in S} \max_{x \in P} d(x,R)$$

The set of points whose distance to a rectangle R is less or equal to ε is called frame of R. The two closed boundaries are called the outer and inner ε-offsets of R. The best fitting rectangle problem is equivalent to computing a narrowest covering ε-frame, the frame with the smallest possible ε that covers all sites. An optimal solution can be found as follows:

- Calculate the major and minor axes of region and rotate region so that both axes are parallel to xy- axes
- Compute the bounding box B of P that results in the outer offset of an optimal ε-frame.
- Compute 2ε = $\max_{x \in P} d(x,B)$
- Let R be the inner ε-offset of B that corresponds to the best fitting rectangle.

Figure 86 shows the different generalized segmentation results for two example buildings. In these cases, rectangle fitting was applied for areas with dormers and chimneys (marked in blue and green). Figure 87 illustrates the generalized superstructures for four example buildings.

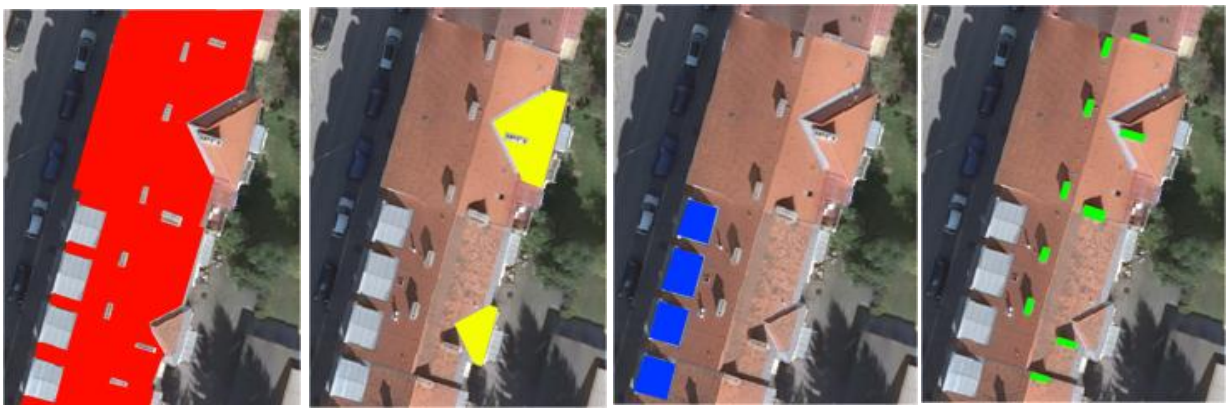


Figure 86: Classification of a test buildings; (a) major roof planes (in red); (b) smaller roof planes (in yellow); (c) dormer windows (in blue); (d) chimneys (in green);

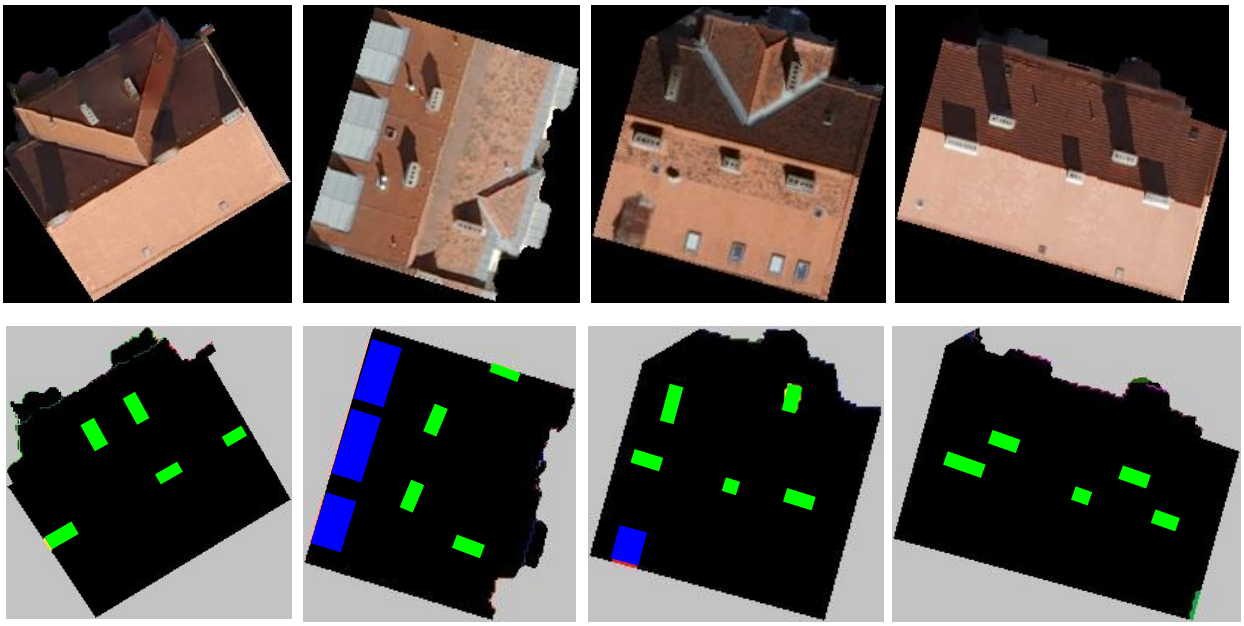


Figure 87: Refined Superstructure Classification after Rectangle Fitting for four example buildings

8.5. Skylight Detection

A major question of roof analysis addresses the use of the attic. The existence of dormer windows strongly indicates that the attic has been extended. However, the absence of dormers does not indicate that the attic has not been extended. In lieu of dormers there might be skylights.

Skylights can be detected by matching a rectangular skylight template with an edge image computed from the aerial photography over the area of the major roof planes. The match is scale and rotation variant. Fast directional chamfer matching as proposed in [Liu et al., 2010]

improves the traditional alignment between two edge images by incorporation of edge information into the matching costs. This improves the robustness of chamfer matching in the presence of background clutter. This method augments the chamfer distance with an additional cost for orientation mismatch, which is given by the average difference in orientations between template edges and their nearest edge points in the query image. Instead of an explicit formulation of the orientation mismatch, the chamfer distance is generalized to points in R^3 for matching directional edge pixels. The directional chamfer distance score is given in [Liu et al., 2010] by

$$d_{DCM}(U,V) = \frac{1}{n}\sum_{u_i \in V} \min_{v_j \in V} \left|u_i - v_j\right| + \lambda\left|\varphi(u_i) - \varphi(v_j)\right|$$

λ … weighting factor between location and orientation terms;

u_i… set of template image edge map

v_j… set of query image edge map

φ()… direction term

Figure 88 shows results for the detection of skylights on building roofs. A visual inspection demonstrates that all skylights of this building roof were successfully detected.

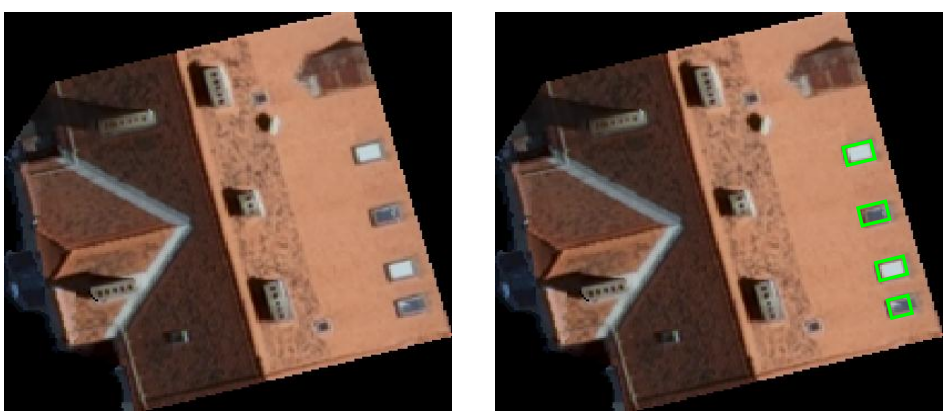


Figure 88: Left: original roof image; Right: detected skylights (marked in green) on a building roof.

8.6. Roof Characterization

The roof planes and the roof classification are now available as an input for a roof characterization. This section is split into two parts: The roof shape of a building and the degree of extension of a rooftop. The analysis is complex because building configurations can be rather complex, thereby making it difficult to characterize them. This fact has an impact on the quality of the roof analysis in dense urban areas and historic areas.

Roof Shapes

To cope with this problem it makes sense to consider the building footprints, which has been done in the modeling of facades in chapter 7. One can decompose a building footprint into so-called "cells" representing simple basic shapes, using a method introduced in [Kada et al., 2009]. Should an individual cell appear to still be complex, we review the 3D point cloud itself. The point cloud of a cell gets associated with different height classes using a method introduced in [Dan, 1996].

It is necessary to differentiate between simple and complex building footprints. The complexity of buildings and roofs may include urban L- and T-shapes (see Figure 89). A simple footprint is of a building consisting of only four faces (see Figure 91). Complexity gets increased if roof elements are connected, as shown in Figure 93: a footprint is thus complex when a building has more than 4 faces and if it has connected roof elements. This is the case for L- ,T- and miscellaneous other buildings like the ones illustrated in Figure 89 and Figure 92. The complexity of the building is thus represented by the complexity of its footprint, and this must be determined.

The big problem is that the building classification in use (see Figure 89) is not very accurate because of misclassifications due to vegetation and shadows. Therefore, to enhance our roof characterization we have to determine the façade walls. We use a method to reconstruct building façades in 3D that was presented in chapter 7. This resulted in a description of the complex 3D façade and a refined building footprint. Figure 89 illustrates how the footprint gets represented by straight line segments. The classification into simple and complex buildings is done as follows:

$$t = \frac{labeled\ pixels\ within\ bounding\ box}{total\ number\ of\ pixels\ within\ bounding\ box}$$

If the threshold t exceeds a certain value, in our case if less than 99% of all pixels belong to the building mask, the building is considered as complex.

Figure 89: (a) building visible in vertical aerial image (b) segmentation result (c) in red modified building outline using 3D façade reconstruction (d) enhanced building footprint.

For our test area, we assume that the majority of buildings have either one section or multiple connected sections. A partition of these sections can be properly derived from the outline polygon. When this is done for each single section a parameterized standard shape can be found. Therefore, we split the footprint into mostly quadrilateral non-intersecting regions to ease a further characterization. A method presented in [Kada et al., 2009] divides a building footprint into mostly quadrilateral shaped polygons in accordance with Figure 90.

The individual cells are created as instances based on a pre-defined set of parameterized cell types. Complex buildings are then modeled in a bottom-up fashion by merging the simple cells.

This can be extremely difficult if the building outline is very detailed and consists of many short line sections. Therefore, instead of using all possible line segments from the outline polygon, an adequate subset must be found. This can be done by defining an adequate subset that still reflects the characteristics of the building. The resulting building outline will be a generalization of the original that resembles the original as close as possible. In this context, a minimum length for each segment has to be defined in advance to keep the number of generated cells low. In our project, this was already considered within the façade footprint determination.

Once the lines have been generated, a rectangle that is two times the minimal bounding rectangle is taken and split by these lines forming non-intersecting cells. The cells are compared to the original footprint, and the ones with a low overlap value are discarded.

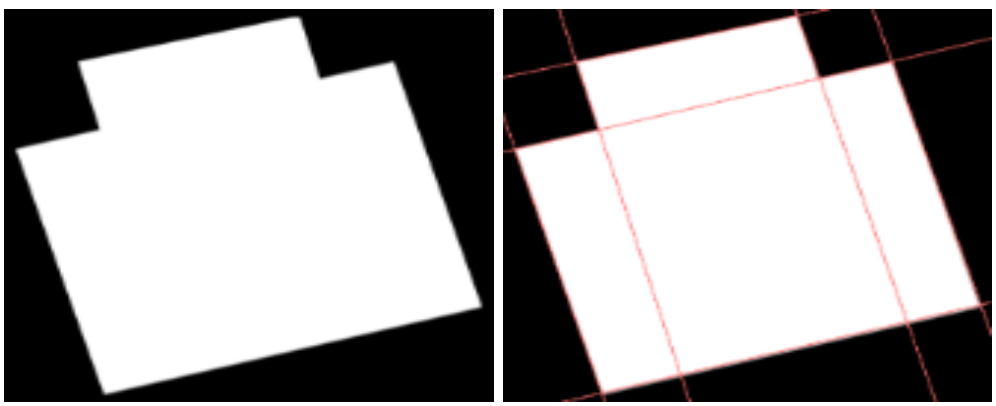


Figure 90: Decomposition of building footprint from figure into quadrilateral cells.

After the building footprint is split into cells, the parameterized roof shapes of all cells can be determined and then combined. We are particularly interested in three types of roofs: The first are standard architectural styles illustrated in Figure 91 and Figure 94, the second are their combinations when roofs get connected as illustrated in Figure 93 and the last are roof shapes that are only available for manual editing.

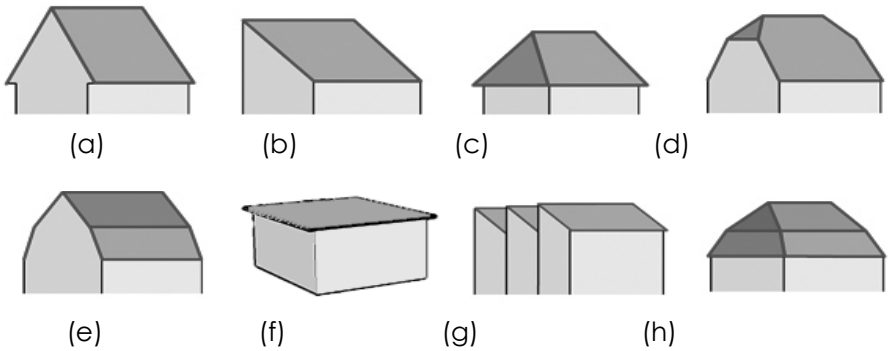


Figure 91: Major architectural roof styles (a) gable roof (b) shed roof (c) hip roof building (d) half hipped roof (e) mansard roof (f) flat roof (g) saw-tooth roof (h) mansard hipped roof. From [www.projekt--service.de/dacharten.html].

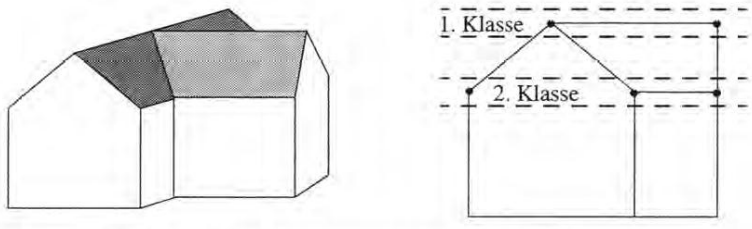


Figure 92: Classification of a building roof into different height classes; Shown are detected lower and upper class for one building roof. taken from [Dan, 1996].

For the determination of the roof type we use a concept introduced by [Dan, 1996] that classifies the 3D point cloud in different height classes (see Figure 92).This idea is very crucial when dealing with different kinds of hip roofs and mansard roofs. In contrast to the method used in[Dan, 1996]we employ both the roof pitch and the building footprint. Therefore all segments of roof planes are assigned to their appropriate height class. The number of points belonging to each of the three classes is counted per plane. Points in the upper and lower class are of special interest. The ratio between the numbers in these two classes is relevant for the assessment of the roof shapes. This is especially effective with hip roofs. The classification uses the following features of a roof:

- Number of planes
- Roof pitch
- Orientation of normal vector
- Points belonging to lower and upper class and their ratio
- Adjacency matrix (planes)

One important parameter is the orientation of the normal vector. This is important to determine the orientation of two adjacent roof planes. The result of this calculation is a function of the position of two roofs; it is 0 if two planes stand normal to each other, 1 if two planes face in a similar direction and -1 if they look in opposite directions. The use of these parameters and the decomposition of a roof into quadrilateral cells results in a classification of each roof into the

major roof shapes of Figure 94. These features are unique for these roof types and enable a proper classification of roof shapes.

In addition to the principal roof types of Figure 94, the approach also works with the connecting roof shapes of Figure 93. One significant improvement of the proposed method over [Kada et al., 2009] is the ability to deal with non-symmetric and more general roof shapes.

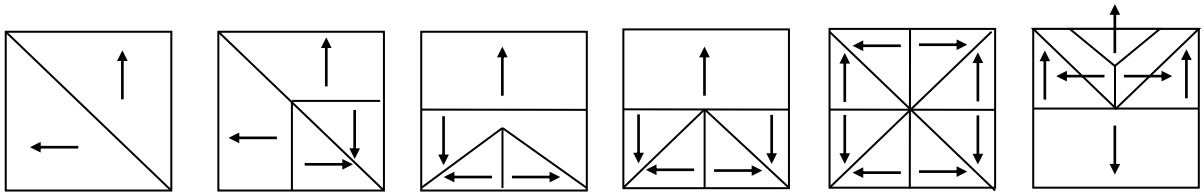


Figure 93. Examples for possible connecting roof shapes with normal vector

Flat roof

Planes	1
Roof Pitch	~ 0 degree
Orientation	-
Classes	-
Ratio	-
Adjacency	-

Half Hipped roof

Planes	4
Roof Pitch	> 0 degree (each 2 are similar)
Orientation	4x is 0; 2x is -1
Classes	2 planes have lower and upper class 2 planes have just upper class
Ratio	upper class smaller than lower class

Shed roof

Planes	1
Roof Pitch	> 0 degree
Orientation	-
Classes	plane has lower, upper class
Ratio	same ratio
Adjacency	-

Mansard roof

Planes	4
Roof Pitch	> 0 degree (each 2 are similar)
Orientation	2x is 1; 1x is -1
Classes	2 planes have upper class 2 planes have lower class
Ratio	- Adjacency 1 neighboring

Gable roof

Planes	2
Roof Pitch	> 0 degree (similar)
Orientation	1x is -1 (dot product of 2 adjacent normal vectors)
Classes	all 2 planes have lower and upper class
Ratio	same ratio
Adjacency	1 neighoring

Mansard Hipped roof

Planes	8
Roof Pitch	> 0 degree (each 4 are similar)
Orientation	4x is 1; 1x is -1; 8x is 0
Classes	4 planes have upper class 4 planes have lower class
Ratio	points in upper class are fewer than in lower class

Hip roof

Planes	4
Roof Pitch	> 0 degree (each 2 are similar)
Orientation	4x is 0; 2x is -1
Classes	4 planes have lower and upper class
Ratio	upper class smaller than lower class
Adjacency	3 planes are neighoring

Sawtooth roof

Planes	several
Roof Pitch	> 0 degree (all are similar)
Orientation	always is 1
Classes	all planes have lower and upper class
Ratio	same ratio
Adjacency	various

Figure 94: Compilation of different features for classification into the roof styles

Rooftop Extensions

Besides the determination of the roof shape of one building, there exists many other questions that can be answered by analyzing vertical aerial images. The most prominent question would be: Is the rooftop of a certain building extended? This is based on the number and location of

chimneys, and the use of skylights and dormers. The question is also interesting for urban planning, real estate development and transactions (it could also be used for value assessments for property tax assessments). Figure 95 illustrates a segmented building roof and based on that the result for the determination of the rooftop extensions.

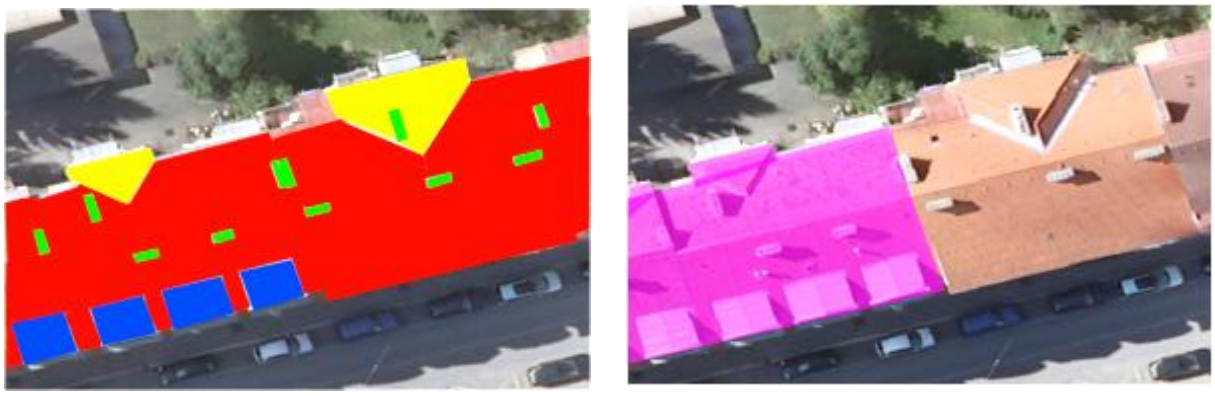


Figure 95: (left) classified roof image; (right) interpreted classification (extended attic is highlighted in pink)

Suitability for the Installation of Solar Panels

Another very prominent issue when dealing with roofs is their suitability for the installation of solar panels. There are numerous publications dealing with this topic, mainly using LIDAR data. We do not research this topic; however, we would like to mention that it is feasible to determine the suitability of a roof using only aerial images. For this purpose, for every roof plane the annual sum of incoming global solar radiation has to be calculated (see Figure 96). More information about this special topic is available at [E.G. JOCHEM ET AL., 2009; KASSNER ET AL., 2008].

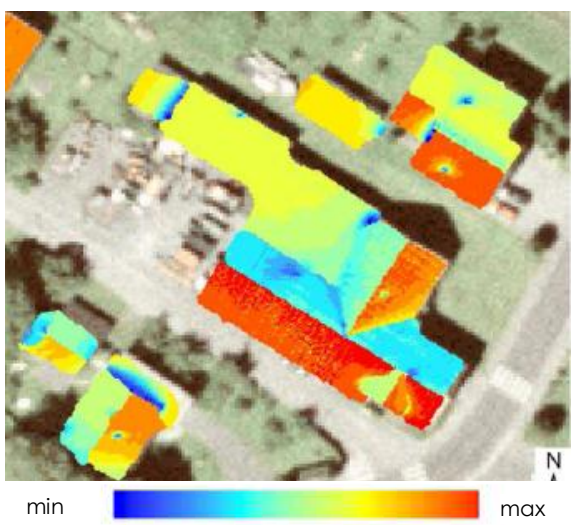


Figure 96: Annual sum of incoming global solar radiation on roof planes in kWh/m2 by considering transparent properties of vegetation; image taken from [JOCHEM ET AL., 2009].

Future work might address the existence of roof terraces, flat roofs with their elevator shafts and air conditioning compressors to characterize the uses of the building.

8.7. Experiments

Within this project, we performed several experiments on roofs using two test data sets, one from Graz, Austria and test data provided from the IGN in Paris covering the dense urban part of the city of Annecy, France. Ground truth of the roof shapes for both data sets was collected by hand. Therefore, building roofs were digitized in 3D using AutoCAD. Figure 97 and Figure 98 illustrate the Ground Truth for the Graz dataset.

Figure 97: Ground Truth for Graz data set (projected on xy-plane) (red: major planes; yellow: smaller planes; blue: dormer windows; green: chimneys) This was collected manually.

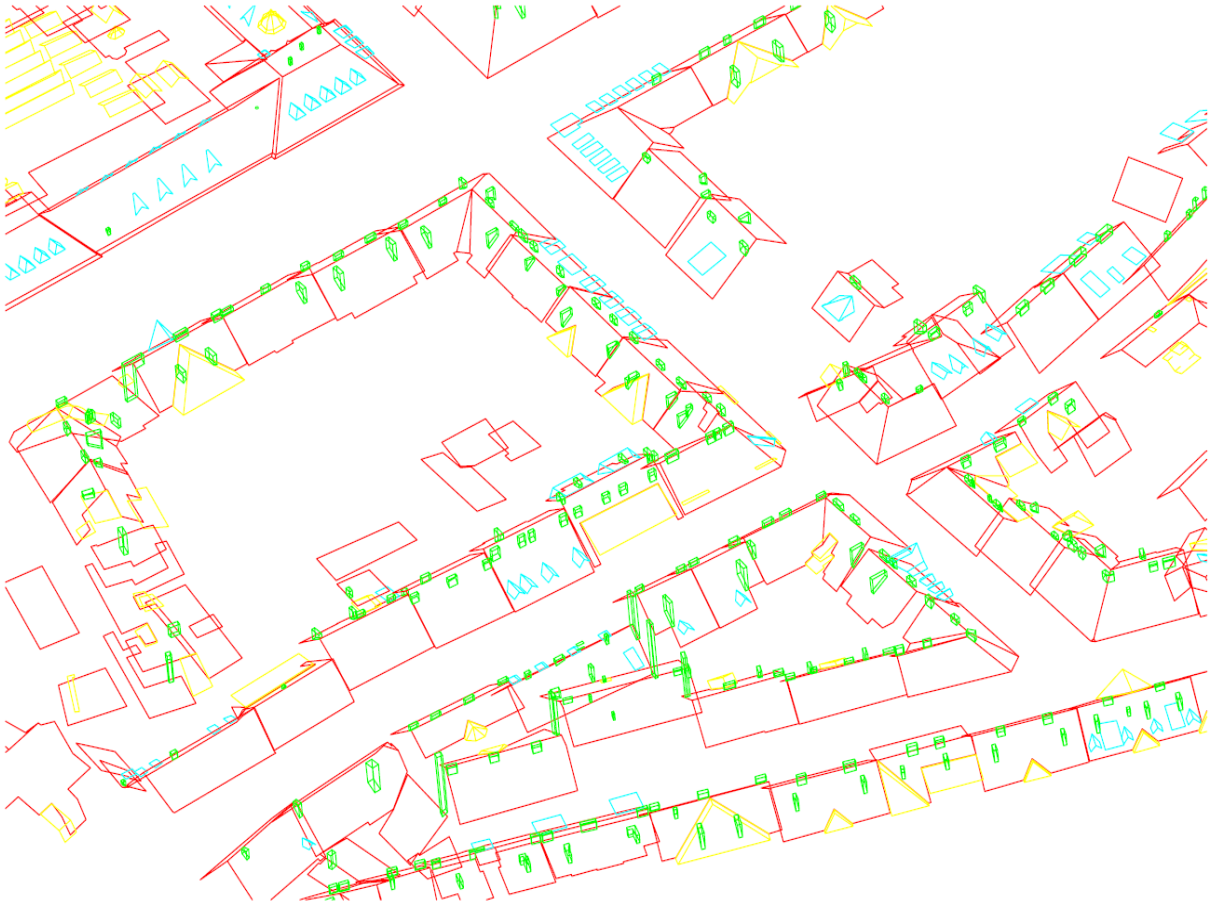


Figure 98: Ground Truth for Graz data set (oblique viewing direction)

DSM Smoothing

In the first experiments, we evaluated the performance of the DSM smoothing step. We applied the TGV^2 model to our Graz data set consisting of 216 buildings. Figure 77 shows the elevation data before and after the smoothing step. The visual inspection shows that inlier noise and outliers are successfully removed and all edges are preserved. However, the smoothing of the DSM not only eliminates outliers, but increases the throughput of plane detection using the method of [Toldo et al., 2008] by a factor of two.

Plane Detection

We have evaluated the plane detection for our Graz data set and the Annecy data set using three different methods. In a first step, we have determined the ground truth for every building of the Graz data set consisting of 216 buildings and for the Annecy data set consisting of 35 buildings. Table 35 summarizes the ground truth for the Graz and for the Annecy data set. For our Graz test area, we determined 512 major planes, 298 smaller planes and 1069 superstructures.

In a first experiment, we have performed RANSAC plane detection for our test data set. In a second experiment, we have used J-Linkage. In a last experiment, we have performed the plane detection using a multi-label segmentation based on label cost minimization.

	Major planes	Smaller planes	Superstructures
Total number (Graz)	512	298	1069
Total number (Annecy)	102	57	429

Table 35: Ground truth of roof planes and roof superstructures for Graz and Annecy data set

Classic RANSAC approach

The classic RANSAC approach for a 3D point cloud for this evaluation. To accelerate the computation and to reduce the number of hypotheses we have used the smoothed point cloud. Table 36 illustrates the result for the classic RANSAC algorithm.

	Major planes		Smaller planes		Superstructures		Total planes	
	Graz	Annecy	Graz	Annecy	Graz	Annecy	Graz	Annecy
Total number	512	102	298	57	1069	429	1879	588
Detected planes	379	74	223	40	737	305	1339	419
Detection rate	74	71	75	71	69	65	71	71

Table 36: Plane detection results using classical RANSAC approach

The application of classic RANSAC algorithm on major planes gives successful results in 71% of the cases that corresponds to the results achieved by [Tarsha-Kurdi et al., 2007a], which achieve detection rates around 70%. Dealing with major planes, we detected 379 out of 512 and 74 out of 102 that corresponds to 74% and 71% for our Graz and Annecy test data sets. For the smaller planes, we detected 223 out of 298 and 40 out of 57 planes for our two data sets. When we look at the results for superstructures, we could determine 737 out of 1069 and 305 out of 429 structures.

The reason why our results for the Graz data set have higher rates than the Annecy data set is based on the origin of the elevation data. The Graz data set is based on high overlapping aerial images that results in a gridwise arrangement of the points. Moreover, the GSD is 10cm. In contrast, the Annecy data is derived from Airborne LiDAR data that has a lower ground sampling distance. Moreover, the elevation data is very noisy. However, in total concerning RANSAC both data sets have the same overall detection rates. The mismatches can be explained by the use of a pure mathematical principle, without taking into account the particularity of the building roof, which is one of the major drawbacks of the RANSAC algorithm.

J-Linkage approach

To improve these results we performed experiments with a method introduced by [Toldo et al., 2008]that incorporates neighborhood relations. Table 37 summarizes the plane detection results using this method.

	Major planes		Smaller planes		Superstructures		Total planes	
	Graz	Annecy	Graz	Annecy	Graz	Annecy	Graz	Annecy
Total number	512	102	298	57	1069	429	1879	588
Detected planes	436	84	244	45	855	330	1535	459
Detection rate	85	82	82	79	80	77	82	78

Table 37: Plane detection results using J-Linkage approach

In contrast to the previous results, this method can increase the detection rate by 11% respectively 7%. We can detect 436 out of 512 major planes that correspond to a detection rate of 85% for the Graz data set and 84 out of the 102 smaller planes that corresponds to a detection rate of 82% for the Annecy data set. Concerning smaller planes, we could detect 244 planes respectively 45 planes that correspond to 82% respectively 79% correct results. 855 out of 1069 superstructures could be successfully detected for the Graz data set and 330 out of 429 could be detected for the Annecy data set.

However, the major problem is that it is mired in a RANSAC-like dependence on manual parameter input. The selection of the inlier threshold and the rejection threshold is crucial for the outcome. Another big problem is that it is not computational efficient when dealing with large urban scenes with millions of data points corrupted by a large amount of gross outliers, which is the case for our data. For these reasons, it is crucial to use a smoothing step before executing this method. Otherwise, the amount of determined hypotheses will be too large and the program will crash.

Multi Label Segmentation approach

In a last experiment, we have evaluated the plane detection using multi-label image segmentation in the context of height model generation. In this case, the input consisted of unsmoothed elevation data. Table 38 summarizes the plane detection results for this method.

As one will notice, we could improve the detection rate by more than 10%. We could detect 481 out of the 512 major planes that correspond to a detection rate of 94% for the Graz data set, and 94 out of 102 planes that correspond to a success rate of 92%. The results for smaller planes are similar, where we could detect 95% for the Graz data set and 90% for the Annecy data set. Concerning superstructures, we could increase out detection rate by 15% for the Graz data set where we could determine 1015 out of 1069 superstructures. For our Annecy data set, we could increase the detection rate from 77% to 86%.

	Major planes		Smaller planes		Superstructures		Total planes	
	Graz	Annecy	Graz	Annecy	Graz	Annecy	Graz	Annecy
Total number	512	102	298	57	1069	429	1879	588
Detected planes	481	94	283	51	1015	368	1779	513
Detection rate	96	92	95	90	95	86	95	87

Table 38: Plane detection results using multi label segmentation approach

Figure 99 shows different height models and segmentation results using different values for λ ranging from 0 to 0.15 and γ ranging from 500 to 1250. The segmentation results provide a meaningful partitioning into the underlying structure of the building.

Figure 100 shows on the left hand side, the input images for five example buildings and the corresponding 3D visualization. Parameters were kept constant at λ = 700 and γ = 0.05. While Figure 99 already shows that a single parameterization can be used for a wide range of input data, we further investigate the influence of the parameter choice. The segmentation results illustrated in Figure 99 show only minor differences, demonstrating that the used method is robust to parameter variations in a reasonable range. When strong regularization is combined with large label costs some regions can disappear.

In Figure 101, we also demonstrate negative examples where our approach fails and where we have a strong over segmentation because of noisy elevation data or non-planar roof shapes. A major limitation for the plane detection presents itself when the roofs are curved or very fragmented, thus when there are no planes. While most regions are successfully detected, some areas are flattened or approximated with several regions. However, one of the major advantages of this method is that the output of the proposed segmentation algorithm is perfectly suited to build a full 3D mesh. The used algorithm was implemented in Matlab and computed on the GPU. The experiments ranged from approximately 10 seconds to 2 minutes per iteration, depending on the complexity and size of the area.

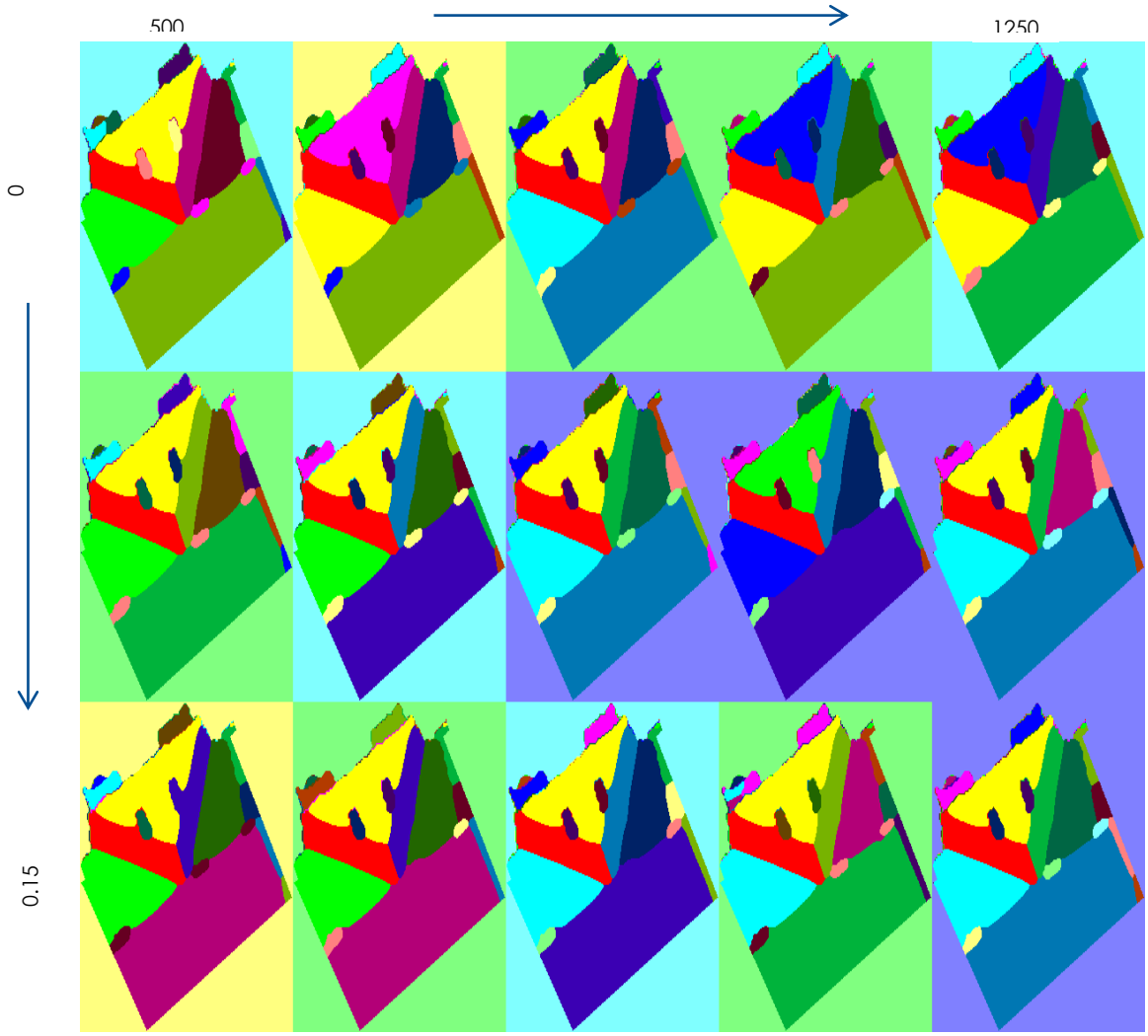


Figure 99: Results for segmentation by variation of the regularization parameters λ and γ. Within a reasonable range, the segmentation result is very robust. values for λ ranging from 0 to 0.15 and γ ranging from 500 to 1250.

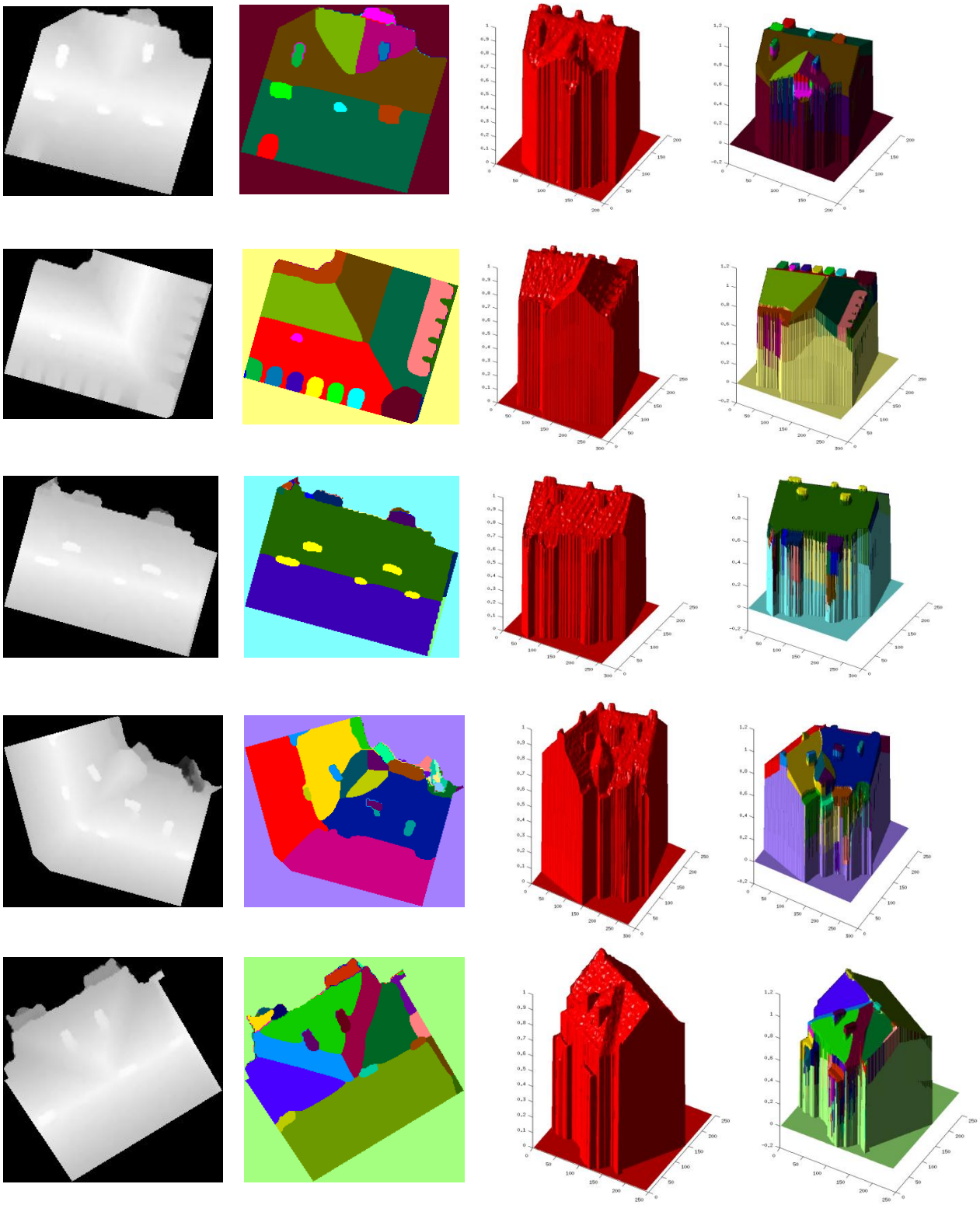


Figure 100: Results for height model segmentation using fixed parameters (λ = 700 and γ = 0.05) (a) Input depth map, segmentation image, input rendering, reconstructed building

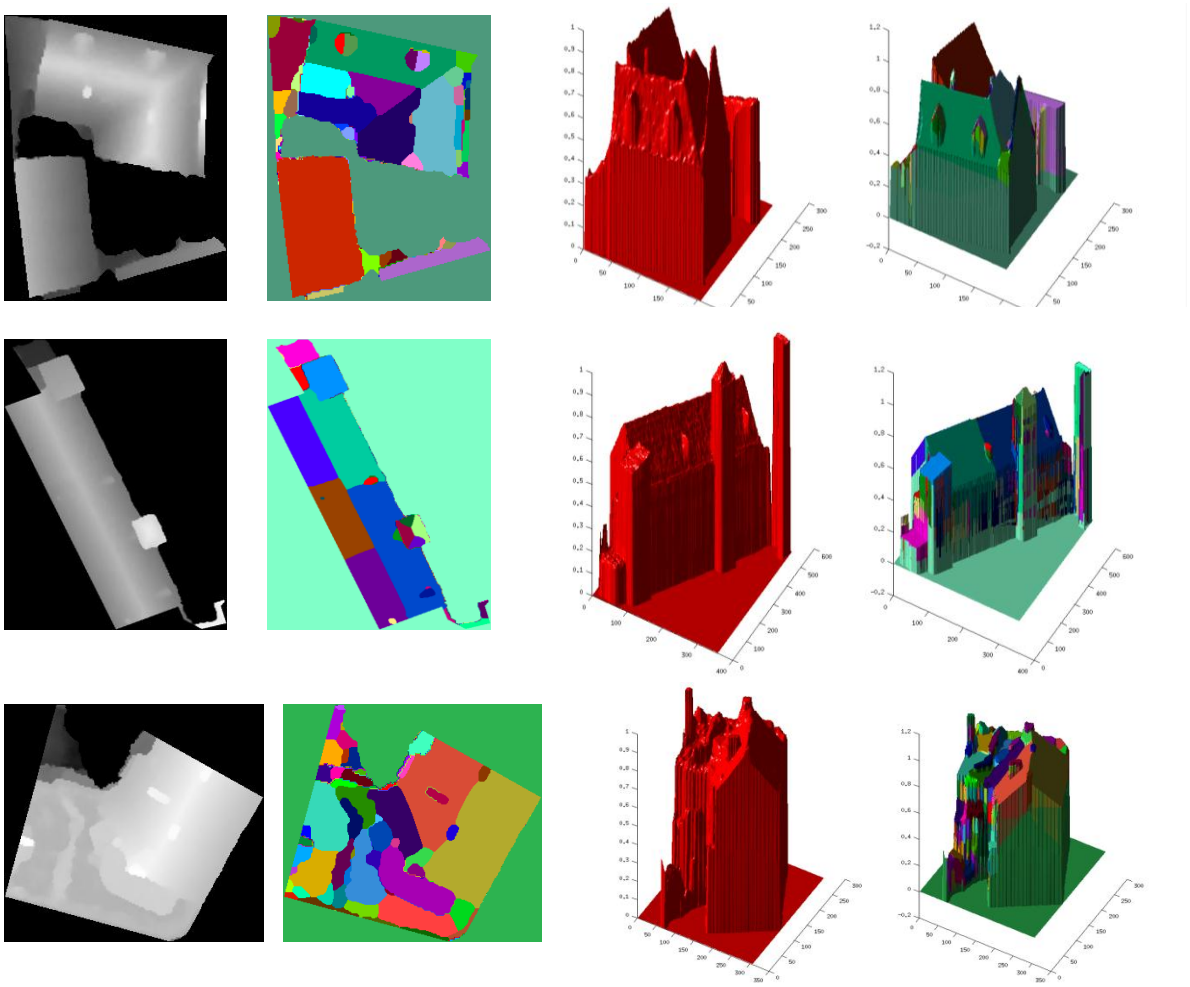


Figure 101: Different examples of errors within the height model segmentation.

Discussion

Current limitations of all these approaches were found in cases where plane segments have very similar pitches and are thus merged erroneously into one plane when they should be kept separate. Another limitation exists with curved roof surfaces that exist in some regions extensively, such as in France, and they may also be found in modern buildings. In these cases, the attempt at fitting roof planes produce many small segments and a failure of the approach. A check will be needed that addresses curved roofs.

Classification

A differentiation between major and minor or smaller planes is heuristic and a function of architectural customs in an area. The threshold can be assigned as a function of every building's footprint and the size of the individual plane segments. But some smaller planes will get erroneously assigned to the major plane class and vice versa. Another category of roof planes are superstructures that are specified by criteria like discontinuities at their edges, size,

shape, location on the roof. These values were determined in a heuristical way and depend on the test area. Before we can classify these planes, we have to collect the ground truth (see Table 39).

Superstructure Classification

The superstructures can be further classified into 225 dormer windows, 586 chimneys and 258 other structures. More than half of the other structures were satellite antennas or walkways on the roofs leading to the different chimneys. For the Annecy data set, the collected superstructures consist of 145 dormer windows, 135 chimneys and 149 other structures.

	Dormer windows	Chimneys	Other structures
Graz	225	586	258
Annecy	145	135	149

Table 39: Ground truth for different types of Superstructures (Dormers, Chimneys and other structures)

For our two test areas, the initial classification was already performed within the plane detection experiments where we divided the planes into three different groups. The results of the different experiments are shown in Table 36, Table 37 and Table 38. In these tables, the detected planes are split into three different groups: major planes, smaller planes and superstructures. In the next part of the experiments, we evaluate the superstructure classification and the refined classification. Therefore, we define that a superstructure is correctly detected if more than 50% of its shape is classified in one of the three categories. We evaluate this processing step using the results of the multi-label segmentation as input data. This method achieved an success rate of 95% for the Graz data set (1015) and 86% for the Annecy data set (368) (see Table 39). Based on these detected superstructures, we perform the further processing steps. Therefore, to evaluate the approach correctly we have to adjust the values for further evaluation (see Table 40).

	Dormer windows	Chimneys	Other structures
Graz	205	564	246
Annecy	130	106	132

Table 40: Ground Truth for Superstructures (Dormers, Chimneys and other structures) for CORRECTLY detected superstructures

The segmentation method missed 20 dormer windows of the Graz data set and 15 dormer windows of the Annecy data set. Concerning the chimneys this method missed 24 respectively 19 chimneys and 12 respectively 17 other structures. Reasons for these mismatches were noisy elevation data and curved surfaces. Especially chimneys on top of the roof were missed by merging the plane belonging to the chimneys with the surrounding planes.

Table 41 and Table 42 summarize the different detection rates for the superstructure classification based on the values from Table 40 for the Graz and the Annecy data set.

	Dormer windows	Chimneys	Other structures
Total number	205	564	246
Detected segments	174	547	206
Detection rate (%)	85	97	84

Table 41: Experimental results for the building roof segmentation showing the detected roof segments (Graz data set)

	Dormer windows	Chimneys	Other structures
Total number	130	106	132
Detected segments	107	101	108
Detection rate (%)	82	95	82

Table 42: Experimental results for the building roof segmentation showing the detected roof segments (Annecy data set)

Dealing with the Graz data set, detected were 174 dormer windows, 547 chimneys and 206 other structures that correspond to a detection rate of 85%, 97% and 84%. For the Annecy data set the superstructure classification resulted in detection rates of 82% for dormer windows, 95% for chimneys and 82% for other structures.

In a next step, we want to evaluate how many dormers, chimneys and other structures were classified in the wrong category. Table 43 and Table 44 illustrate the analysis of the occurring mismatches for both data sets.

	Dormer windows	Chimneys	Other structures	m./s. planes
Dormer windows	174	6	11	14
Chimneys	5	547	8	4
Other structures	20	13	206	7

Table 43: Misclassification of superstructures (Graz data set)

	Dormer windows	Chimneys	Other structures	m./s. planes
Dormer windows	107	4	10	9
Chimneys	0	101	4	1
Other structures	6	13	108	5

Table 44: Misclassification of superstructures (Annecy data set)

For the Graz test data set, 6 dormer windows were classified as chimneys and 11 were classified as other structures. The results for the Annecy data set have similar values for misclassifications. Concerning chimneys 8 respectively 4 were classified as other structures and 4 were classified as dormer windows. 20 other structures were classified as dormer windows and 13 were classified as chimneys. For the Annecy data set, 6 other structures were classified as dormer windows and 13 were classified as chimneys. Of special interest are the values in the last columns of table 43 and table 44. 23 dormer windows, 5 chimneys and 12 other structures were classified as major smaller planes. In these cases, the structures were classified in the same category as their surrounding regions. This happens when structures are not elevated enough or when they are too small to be detected. In the case of dormer windows, this misclassification happens due to their large size and position on the roof.

The shape of other structures causes problems within the classification process. For instance, trees on roof terraces or satellite antennas are often detected as chimneys because of their similar shapes. A possible solution could be the incorporation of CIR (Color Infrared) images in the segmentation process to reduce the misclassification caused by vegetation.

Beside these misclassifications, within the experiments, we also had to deal with false positives. Table 45 demonstrates the false positives for our two data sets. For example, 9 chimneys in the Graz data set and 12 chimneys in the Annecy data set were detected by mistake. Reasons for these wrong detections are outliers within the elevation data and complex roof shapes.

	Dormer windows	Chimneys	Other structures
Graz dataset	2	9	15
Annecy dataset	1	12	24

Table 45: Superstructure Detection: False positives

Refined Classification

In the next part of the experiments, we want to determine the influence of the region growing for the classification. Therefore, we have overlaid the ground truth with the segmentation result and have determined the percentage of correctly classified pixels before and after the region growing. Table 46 and Table 47 demonstrate the classification results before region growing, and Table 48 and Table 49 illustrate the results after this refinement step.

	Dormer windows	Chimneys	Other structures
Correct detected	77%	84%	78%
Wrong detected	23%	16%	22%

Table 46: Accuracy analysis of classification step before region growing (Graz data set)

	Dormer windows	Chimneys	Other structures
Correct detected	79%	85%	72%
Wrong detected	21%	15%	28%

Table 47: Accuracy analysis of classification step before region growing (Annecy data set)

	Dormer windows	Chimneys	Other structures
Correct detected	85%	97%	84%
Wrong detected	15%	3%	16%

Table 48: Accuracy analysis of classification step after region growing (Annecy data set)

	Dormer windows	Chimneys	Other structures
Correct detected	87%	96%	80%
Wrong detected	13%	4%	20%

Table 49: Accuracy analysis of classification step after region growing (Annecy data set)

The analysis of the Graz data set leads to the following results: Dealing with dormer windows, 77% of the pixels belonging to this category were correctly assigned, for chimneys 84% of all pixels were correctly assigned and for the category other structures 78% of all pixels were correctly assigned. Concerning the Annecy data set, 79% of the pixels belonging to this category were correctly assigned, 85% of the chimney pixels were correctly assigned and 84% of the pixels belonging to other structures were correctly assigned.

After region growing we could increase these accuracies by several percent. For chimneys we could increase the values from 84% to 97% respectively from 85% to 96%. This is due to the well

defined edges. The accuracy of the classification for dormer windows and other structures could also be increased, although the increase is not as high as for chimneys.

Discussion

Misclassifications of superstructures occurred in the case of very complex roof shapes with non-allocable roof structures, particularly when surfaces were curved. Since our approach had not been considering the existence of roof gardens, it failed in those cases. Thus, it is obvious that further work is needed to address curved roofs and roof terraces.

Skylight detection

In a next step, we want to evaluate the accuracy of the skylight detection. Therefore, we have manually collected the ground truth for the Graz data set (see Table 50). For the detection, we have used roof regions assigned as smaller and major planes. For the 764 successfully detected major and smaller planes, we generated edge images using Canny edge detection and converted them into distance images. Using these distance images and a binary template of a window, in our case a rectangle, we perform the chamfer matching. Table 50 illustrates the results of this processing step. We successfully detected 453 out of 574 skylights, which corresponded to a detection rate of 79%. Most of the 189 misclassifications occurred due to rectangular patterns on the roof resembling a window.

	Skylights
Total number	574
Detected segments	453
False positives	189
Detection rate (%)	79

Table 50: Detection rates for skylight detection for Graz data set

Roof characterization

In the next part of this project, we evaluated two major topics. In a first step, we analyze the accuracy of the roof shape determination. In a second step, we analyze the accuracy of the rooftop extension determination. In both cases, we use examples from the Graz data set.

Roof Shape

Table 51 summarizes the ground truth of roof shapes for the Graz data set. Since some roofs fall outside the standard types, a "miscellaneous" category with 24 entries was required. An additional classification was into "simple" and "complex" buildings, and the latter was further classified into L-buildings (corners), T-buildings and buildings with U- and other irregular shapes. The test area has 147 simple and 69 complex buildings. Of the complex buildings, 25 are L-and 17 are T-buildings, and 19 fall into the miscellaneous category.

Flat Roof	Shed Roof	Gable Roof	Hip Roof	Half Hipped Roof	Mansard Roof	Mansard Hipped Roof	Sawtooth Roof	Non-Allocatable Roofs
25	19	123	10	13	0	0	2	24

Table 51; Ground truth: types of roof shapes for the 216-building test area

Down-sampling of the DSM point clouds by a factor of 4 is acceptable to suppress minuscule plane segments and to accelerate the throughput. This down-sampling eliminates small structures in advance. Nonetheless, even with this down-sampling not all small structures can be eliminated, especially larger dormer windows. Before the roof type can be assigned, one needs to eliminate the remaining small structures caused by bigger dormers and chimneys and other extrusions or intrusions. Success in defining and removing such small structures was at 98%, 1047 of the 1069 smaller roof structures, and 94% detection rate for the detection of major and smaller roof planes (see Table 52). Without down-sampling we achieved a detection rate of 95% for superstructures that demonstrates a small but crucial increase in accuracy for superstructure detection. Concerning roof planes the detection rate was stable.

	roof planes	superstructures
total	810	1069
detected	765	1047
Detection rate [%]	94	98

Table 52: Detection rate for roof planes and superstructures after downsampling the range data by a factor four.

Before we can evaluate the roof shapes in a first experiment, we evaluated the footprint extraction using the plane sweep described in chapter 7.Figure 102 demonstrates the building footprints based on the image classification and after incorporating information coming from the 3D façade reconstruction.

Figure 102: Building footprints based on classification image and after incorporating 3D façade information

Figure 103 illustrates the differences between both footprints. The green lines represent the original building footprints and the red lines the new ones. A visual inspection shows that all building footprints were correctly determined. It also demonstrates that even with the achieved accuracy of 89% for building classification, there are big mismatches at the borders of the buildings. These mismatches are caused by errors within the image classification due to vegetation and roof and building overhangs.

Figure 103: Overlay of building footprints from Figure 102. Detail of overlay shows discrepancies at building edges due to errors within the image classification. Building footprint before enhancement (red) and derived footprint after 3D reconstruction of building facades (green)

After the decomposition the 216 building from the Graz test data set are split into 827 cells. Figure 104 illustrates the result of the cell decomposition for a small area of the test area. We assign each of these cells to one of the roof categories listed in Figure 93 and Figure 94.

Figure 104: Cell Decomposition. Buildings are split into rectangular areas for further processing.

Then in a next step all the analyzed cells belonging to one building are combined and a common roof type is associated to these buildings. Table 53 shows the detection results for these cells.

Roof shape	Flat roof	Shed roof	Gable roof	Hip roof	Half Hipped Roof	Mansard roof	Mansard Hipped roof	Non allocatable shapes	connected	Saw-tooth roof	Total
Total number	101	98	401	58	25	5	0	25	158	2	873
Detected roof shapes	93	87	368	42	15	4	0	0	135	2	746
Detection rate (%)	92	89	92	72	60	80	-	0	85	100	85

Table 53: Evaluation of roof shapes based on decomposed cells (Result including all cells)

The method could successfully detect 93 out of 101 flat roofs, 87 out of 98 shade roofs, 368 out of 401 gable roofs, 42 out of 58 hip roofs, 15 out of 25 half hipped roofs, 4 out of 5 mansard roofs and all saw-tooth roofs corresponding to an overall detection rate of 85%. However, there are two categories of major interest in this table. There were 25 roof parts with non-allocatable

shapes and 158 connected roof shapes. The non-allocatable shapes belong mostly to modern buildings with very complex roof structures (see Figure 105). These buildings have to be interpreted manually due to the complexity of their roofs.

Figure 105: Examples for non allocatable building roofs (Graz data set showing Graz Opera House and extended rooftop of modern residential building)

After determining the roof shapes for all roof cells in a next step, we combine these cells belonging to the appropriate buildings. Therefore, the results of all cells belonging to one building are collected and their area and shape is analyzed. We explain this merging for one example building. An example building was decomposed into 4 cells. These cells were assigned to the following categories: 2x gable roof, 1 connected roof shape and 1 shed roof. In a next step, we look at the size of these cells. Gable roofs cover 50% of the roof, the connected roof shape 40% and the shed roof 10%. For the characterization, all connected roof shapes were neglected and just the standard shapes were used for the determination of the final shape. In this case, the biggest structures are the gable roofs and therefore the whole building is assigned as a building with a gable roof shape.

The assignment of a standard roof type concerns 192 of the 216 buildings, since 24 fall outside the standards. If we could detect these non-allocatable buildings automatically we achieve a detection rate of 88%, thus 169 buildings out of 192 were correctly classified (see Table 54). If we include the 24 non-allocatable buildings we achieve 78%.

Roof shape	Flat roof	Shed roof	Gable roof	Hip roof	Half Hipped Roof	Mansard roof	Mansard Hipped roof	Saw-tooth roof	Total
Total number	25	19	123	10	13	0	0	2	192
Detected roof shapes	22	16	115	8	6	0	0	2	169
Detection rate (%)	88	86	93	80	43	-	-	100	78

Table 54. Evaluation of roof shape detection for Graz data set

While the flat roof would seem to be the easiest type to classify, only 88% were correctly identified. This translates to an error in 3 of 25 buildings. Additionally, 3 out of the 25 flat roofs had gardens that caused major problems when characterizing roofs. The shed roofs were identified at a rate of 86%, since 3 buildings were incorrectly classified due to large extrusions in the form of large dormer windows. Gable roofs were detected at a 93% success rate. Problems again occur with large extrusions. 8 out of 10 hip roofs were correctly detected. However, the half-hipped roofs were only correctly classified in 43% of the cases. The problem with this category derives from the two small plane segments in the upper height class. The current approach eliminates those plane segments in most of the cases because of the down-sampling of the point density, thus making it undetectable.

The test area did not include any mansard or mansard hipped roofs. We therefore processed 20 additional buildings from another data set (Annecy, France) for each of these two types. Table 55 illustrates the results for these two roof types.

Roof shape	Mansard roof	Mansard Hipped roof	Total
Total number	12	8	20
Detected roof shapes	10	4	14
Detection rate (%)	83	50	70

Table 55: Evaluation of roof shape detection for Annecy data set

10 out of 12 mansard roofs were detected. Errors occurred with the mansard hipped roofs in 4 of the 8 cases. The plane segments in such roofs are at very similar pitches so that plane detection merges plane segments when they should be kept separate. Additionally, old roofs that are not planar caused problems.

Rooftop Extensions

Ground truth was obtained by a manual interpretation of 69 extended attics in the 216 buildings from the Graz data set (see Table 56). The extended attics are characterized in 19 instances by dormer windows, in 27 instances with just skylights and in 23 instances with both dormers as well as skylights. Figure 106 illustrates the ground truth for the Graz data set. Buildings marked in yellow only have dormer windows, red buildings only have skylights and green buildings have both. By incorporating the results achieved within the classification step and the template matching, we could successfully detect 60 out of 69 extended attics that corresponds to a detection rate of 87%. In 10 cases, our method wrongly detected extended rooftops. Reasons for these errors are complex roof shapes and errors within the skylight detection.

	Dormer windows	skylights	both	total
number	19	27	23	69

Table 56: Ground truth for 69 extended attics in 216 buildings. Some attics have only dormer windows or skylights, and some have both.

Figure 107 demonstrates the results of the rooftop extension analysis for the Graz test area. Correctly detected buildings with an extended rooftop are highlighted in green, missed rooftops are in red and false negatives are in blue. This analysis about the degree of extensions of rooftops is especially interesting for urban planning and investors because it strongly effects the value of a building. With this method, we can determine the potential of the rooftops for a certain area within a short amount of time.

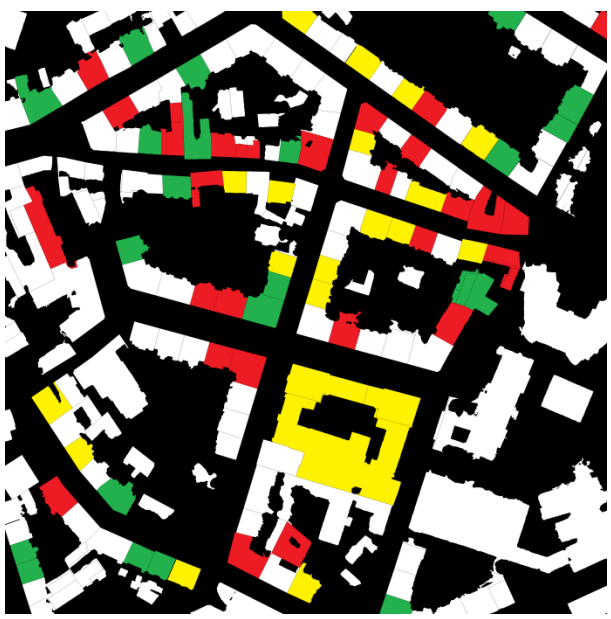


Figure 106: Ground truth for Graz test data set. Yellow: Buildings with dormer windows; Red: Buildings with skylights; Green: Buildings with dormer windows and skylights.

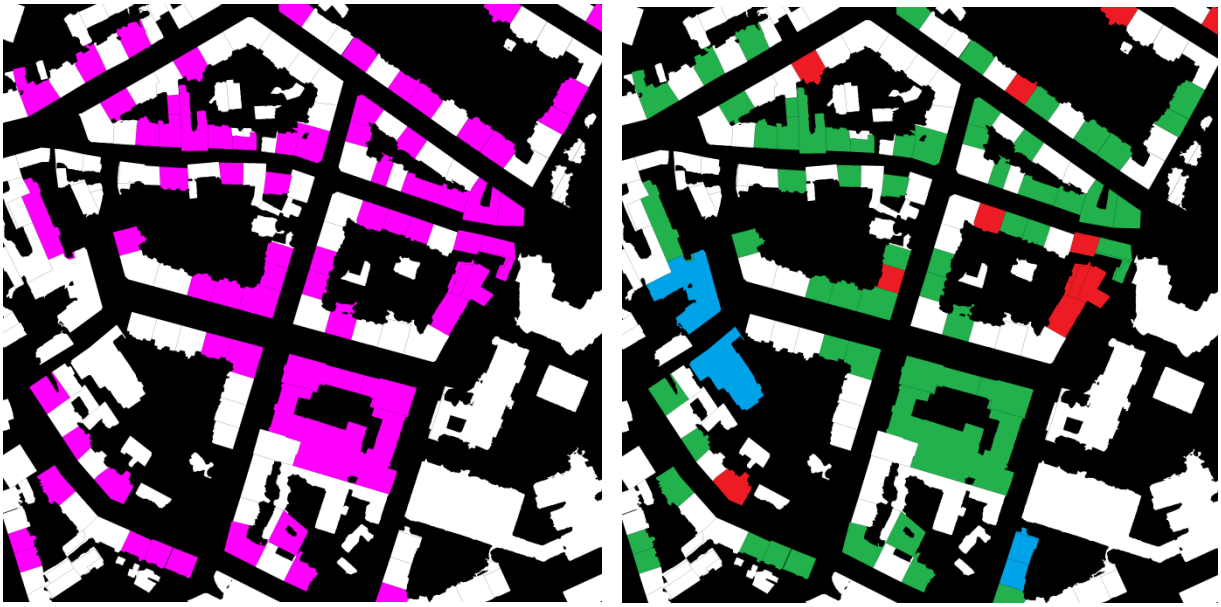


Figure 107: Results of Rooftop Analysis. (left) Ground Truth for extended building rooftops. Extended rooftops are highlighted in pink; (right) Result of Rooftop Analysis. (green) correctly detected extended rooftops (red) false negatives (blue) false positives

8.8. Discussion

We contribute a novel use of vertical aerial images to describe and characterize building roofs as a new content in a 3D GIS. Since the aerial photographs offer a very high overlap, they produce very dense 3D point clouds. These in turn support, jointly with the RGB-content of the images, the mapping of major roof planes and of roof superstructures. The rate of correctly detecting roofs, dormer windows, chimneys and skylights are above 90%. This in turns permits one to assign to a "normal" building a label of "extended attic" and assign specific architectural roof styles.

Moreover, we propose a method for automatically mapping roofs and classifying them into architecturally accepted standard types, based on traditional digital large format color aerial photography. This method relies on point clouds at 100 pixel/m^2 extracted from highly overlapping vertical aerial imagery, and on an image classification using color and texture to delineate building outlines and footprints. Experimental work in two test areas results in correct roof planes in 94% of the cases. Small roof structures do confuse the analysis and must therefore be detected and eliminated. This is successful in 98% of all the test cases. Roof types get classified correctly at a rate of 88%. LiDAR appears as the technology alternative to the use of aerial photography. However, a review on LiDAR literature indicates that there are no statements about the reliability of roof shape detection.

Limitations exist with complex roof shapes that include large dormers, or with curved roofs or if a roof is too complex. However, those difficulties are not specific to aerial photography and also present themselves with LiDAR data. Thus, we conclude that aerial photography produces results at least as good as those from LiDAR point clouds. One of the biggest limitations of assigning architectural roof styles is that sometimes there is no "style" and human perception fails to determine the appropriate roof. This is the case for 24 buildings of our test area where a single roof consists of multiple shapes.

LiDAR today is the workhorse for the production of 3D point clouds, and research in building reconstruction is often considering LiDAR as the major data source. However, a search for LiDAR-literature addressing roof superstructures came up empty. We will therefore argue that vertical aerial photography is a valid data source for mapping building roofs, and that the same cannot be said at this time for LiDAR.

While we propose to continue with the development of roof analysis work based on aerial photography, the experimental work has revealed that weaknesses do exist in the proposed method. These weaknesses from confusions with large dormers, or roof gardens, or small differences in pitch angles of mansard hipped roofs should be addressed. Additionally, we expect that a focus will be needed on the analysis of the detailed extrusions, sometimes intrusions, on roofs. These concern chimneys, dormers, sky lights, terraces etc. Success with those details will reflect back on an improvement of the roof analysis and type assignments.

CHAPTER NINE

CONCLUSION AND OUTLOOK

9.1. Building Modelling

At issue in this project is the motivation for the development of descriptive data about real properties (or real estate) to ease real estate valuation. An initial driver to embark on this research has been the disappointment with the lack of detailed information about objects shown in Internet-maps on Google Earth or Bing/Maps, and the expectation that such information could be available at very little extra cost from image sources in existence for other purposes. Were such detail information to exist, it would be useful in location-based decisions addressing both commercial and public interest environments. Such information would also be useful for a more efficient administration of municipal resources.

9.2. Building Characterization

This project addressed the problem of semantic image interpretation using vertical aerial images, utilizing high redundancy, appearance and 3D information. These high redundancies resulting from the highly overlapping images make such a characterization of single buildings feasible with automated methods.

We delineate how the transition from a characterization to a valuation could be feasible but focus only on the characterization. Valuation is the vision, characterization is the prerequisite. Within this work, we only use vertical aerial images. The aim is to determine the accuracy and completeness of such a characterization. We have to cope with many challenges: there exists a number of different data sources coming from different sensors. Each of these sensors offers data with different accuracies. The main issue we were confronted with was: which of the sensors is suited to solve this task. Are vertical aerial images suited to overcome all posing problems or are steep viewing angles prohibiting their use? When dealing with building roofs, it is obvious that aerial imagery or LiDAR can be suited to meet the expectations. However, when dealing with building facades, the problem gets far more complex. Many other imaging techniques, like street side imagery, are likely better suited for this purpose. We show that facades are visible using vertical aerial images. Moreover, they have another main advantage

compared to other data sources, which is that they already exist for most urban parts of the world, are free of charge and are updated in regular intervals.

When we talk about characterization, there are three main topics of interest: manmade structures like buildings, other structures like impervious surfaces, vegetation and water and third the interaction between buildings and other structures. This project focuses on buildings only. The other areas were beyond the scope of this project. We spent a certain limited effort in assessing the benefits of aerial imagery vs. LiDAR, and came to the conclusion that for the characterization of buildings, aerial imagery offers more building details than LiDAR due to higher point densities.

We also spent a certain effort on comparing oblique vs. vertical photography, and arrived at the conclusion that vertical aerial images are superior to oblique imagery because of the higher resolution and less occlusions in dense urban areas. Our evaluation of the floor and window detection resulted in detection rates of 93% and 85% for planar facades and 86% and 80% for complex roofs. In contrast, by using oblique aerial images the detection rate for floors is around 75% and 66% for windows. Depending on the number of used input images, façades could be reconstructed from vertical aerial images with accuracy between ±13cm to ±19cm at GSD of 10cm.

Buildings are a very complex research topic. When dealing with buildings, one wants to determine all details visible from the outside. Does a building have a cellar, is the rooftop extended, how old is the building, how many floors does the building have, how many windows does a building have and how large are they, are there balconies, what is the building footprint without roof overhangs, are there superstructures on the rooftop, etc.? Of course, we also would like to determine details from the interior of a building. An important aspect are the number of floors and the floor area. However, we cannot give details about rooms inside a building, such as ceilings and furniture, because we cannot see through walls. All these parameters make this project a very complex computer vision topic.

9.3. Accuracies

We achieve detection rates between 75 and 98% depending on the building characteristic. The question now becomes: Is the extracted information useful? Depending on the application field, the requirements differ. Is the success rate, as demonstrated, useful? This would require that the transition of characterization to valuation be implemented. We can only speculate at this time that an 80% success rate will be ample sufficient to approximate a valuation within 20 percent. Therefore, for taxation purposes, we believe that the achieved results should be accurate enough to ease the necessary valuation. However, the uncertainties of such a valuation are enormous. The question if a building is worth one million euro or more depends always on the market.

9.4. Diversity of Data Sources

Nevertheless, we believe that vertical aerial photography, with high overlaps and GSD below 8 cm, is a very useful tool, and that there is no need for neither LiDAR nor oblique imagery. In this work, we showed that vertical aerial images can cover all their topics in connection with the modeling and semantic interpretation of dense urban cores.

The more thorough use of vertical aerial photography is globally expected to get augmented by significant efforts to conflate various 2D Geodata sources, to add in business and private address data bases and parcel data, and by the start of massive collections of GNSS and cellular traces. This is separate from an increasing interest in human scale urban data that is reflected initially in the form of street-side imagery taken by moving vehicles, but in the long run is expected to come from Community-Photo-Collections CPCs, such as those in FLICKR. This in turn will be part of a broader concept of user-generated content UGC.

9.5. Next Steps in Continued Research

Major open issues exist: The complexity of describing buildings is great so that much work remains to be done. It is important to obtain a good understanding of the accuracy and reliability of a building's key numbers. One needs to solve the problems one will have when parts of a building are occluded, when the geometric resolution of the source data varies, when buildings deviate from expected shapes in the event of add-ons and of complex footprints and roof shapes, when the building's use is mixed and an industrial use leads to windowless facades, when cadastral detail contradicts image detail etc.

Lot of work still has to be done to be able to use such a characterization for property valuation purposes. During my research I focused especially on buildings, however there are several open topics left for further research. Especially vegetation and impervious surfaces are of interest. Another interesting and important issue is the interaction of properties like regarding views from windows and occlusions. "Can I see the Schlossberg from my flat? How far can I see from my windows? Which building is blocking my view to see the Schlossberg? Which flats in Vienna are for rent with a view to St. Stephans Cathedral? etc." Concerning occlusions open questions are: "How much sun has this flat? Can I install solar panels on my roof or on my façade?" All these questions can be answered just using vertical aerial images and has to be topic of further research.

The Vision for 2020?

Much of Internet-mapping is currently a result of the search application. However, one may be able to see on the horizon an advent of spatial information as an integral part of the evolution of the Internet-of-Things ("IoT") and of Ambient Intelligence ("AmI"), as projected by [Weiser.,1991] or [O'Reilly et al., 2009].

We expect that by 2020 all these open issues and questions will be successfully solved by ongoing research. Property valuation using vertical aerial images could prevent another financial crisis as the one in 2008 by valuating buildings depending on objective parameter and in real time. This would mean that mortgages can be calculated worldwide using a common scheme that's more transparent! This would lead to a greater safety for banks and insurance companies.

Besides "valuation" there is the vision of detailed 3D urban data being available everywhere, at all times, via smart phones in everyone's hands. This will require the evolution of community photo collections and their automated analysis for 3D modeling and it will require the intensification of location-ware applications like the smart phone.

BIBLIOGRAPHY

[Abellanas et al., 2004] Abellanas, M., Hurtado, F., Icking, C., Ma, L., Palop, B. and Ramos, P.A. 2004. Best Fitting Rectangles.Technical Report 322, FernUniversität Hagen, Department of Computer Science, Germany.

[Agarwal et al., 2009] Agarwal, S., Snavely, N., Simon, I., Seitz, S. and Szeliski, R. 2009. *Building Rome in a Day*.In: In: Proceedings IEEE Computer Society International Conference on In Computer Vision, Kyoto, Japan, pp. 72-79.

[Alegre et al., 2004] Alegre, F. and Dellaert, F. 2004. *A Probabilistic Approach to the Semantic Interpretation of Building Facades*. In: Proc. International Workshop on Vision Techniques Applied to the Rehabilitation of City Centres, Lisbon, Portugal.

[Alharthy et al., 2002] Alharthy, A., Bethel, J. 2002. *Heuristic filtering and 3d feature extraction from LIDAR data*. *ISPRS* Commission III, Graz, Austria.

[Alt et al., 2001] Alt, H. and Hurtado, F. 2001. *Packing convex polygons into rectangular boxes*. In: Discrete and Computational Geometry—Japanese Conference, JCDCG 200, vol. 2098, in: Lecture Notes in Computer Science, Springer, Berlin, Germany, pp. 67–80.

[Ameri, 2000] Ameri, B. 2000. *Automatic Recognition and 3D Reconstruction of Buildings through Computer Vision and Digital Photogrammetry*.Deutsche Geodätische Komission, Reihe C, Nr. 526, München.

[Anders, 2005] Anders, K.-H. 2005. *Level of Detail Generation of 3D Building Groups by Aggregation and Typification*. In: Proceedings of the XXII International Cartographic Conference. La Coruna, Spain.

[ASPRS 1994] ASPRS, 1994. *Mapping & Remote Sensing Tools for the 21st century*. Washington, D.C., USA. 248 pages.

[ASPRS 2009] ASPRS, 2009. *Photogrammetric Engineering and Remote Sensing*. Published by the American Society for Photogrammetry and Remote Sensing ASPRS, Baltimore, USA.

[Avery et al., 2001] Avery T.E. and Burkhart H. 2001.*Forest Measurements*. McGraw-Hill Science/Engineering/Math; 5 edition

[Baillard et al., 1999a] Baillard, C. and Zisserman, A. 1999. *Automatic reconstruction of piecewise planar models from multiple views*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol 2, pp. 559-565.

[Baillard et al., 1999b] Baillard, C., Schmid, C., Zisserman, A. and Fitzgibbon, A., 1999. *Automatic line matching and 3D reconstruction of buildings from multiple views*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 32, Part 3-2W5 "Automatic Extraction of GIS Objects from Digital Imagery", Munich, September, pp. 69-80.

[Baillard et al., 2000] Baillard, C. and Zisserman, A. 1999. A plane-sweep strategy for the 3D reconstruction of buildings from multiple images. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 33, (Part B2), pp. 56-62, Amsterdam, Netherlands.

[Banno et al., 2008] Banno, A., Masuda, T., Oishi, T., and Ikeuchi, K. 2008. *Flying laser range sensor for large-scale site-modeling and its applications in bayon digital archival project*. In: International Journal of Computer Vision, vol. 78, no. (2-3).

[BC-Carms, 2006] Centre for Applied Remote Sensing and Simulation 2006. LiDAR – Overview of Technology, Applications, Market Features & Industry. Technical Report, University of Victora.P. 49.

[Becker et al., 2008] Becker, S. and Haala, N. 2008. *Integrated LiDAR and image processing for the modeling of building façades*. In: Photogrammetrie, Fernerkundung, Geoinformation (PFG), 2/2008, pp. 65-81.

[Becker, 2009] Becker, S. 2009. *Generation and application of rules for quality dependent facade reconstruction* In: ISPRS Journal of Photogrammetry and Remote Sensing, vol. 64, pp. 640-653.

[Becker et al., 2009] Becker, S. and Haala, N. 2009. *Quality Dependent Reconstruction of Building Façades*. In: QuaCon pp. 174-184.

[Berthod et al., 1995] Berthod, M., Gabet, L., Giraudon, G. and Lotti, J. 1995. *High Resolution Stereo for the Detection of Buildings*. In: *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Virkhauser Verlag, Basel, Switzerland, pp. 135-144.

[Binford, 1971] Binford, T.O. 1971. *Visual Perception by Computers*.Invited Paper IEEE Systems Science and Cybernetics; Miami, USA.

[Binford, 1976a] Binford T.O. 1976a. Presentation of first DARPA Image Understanding Workshop, Washington DC, USA.

[Binford, 1976b] Binford T.O. 1976b. Presentation of second DARPA Image Understanding Workshop, University of Md, USA.

[Binford et al., 1977] Binford T.O., Arnold, R. and Gennery, D. 1977. *Recent Research in Computer Science*. In: *Proceedings of DARPA Image Understanding Workshop*.

[Birchfeld et al., 1998] Birchfield, S. and Tomasi, C. 1998. *Depth discontinuities by pixel to- pixel stereo*.In: Proceedings IEEE Computer Society International Conference on In Computer Vision.

[Bleyer et al., 2010] Bleyer, M. and Chambon, S. 2010. *Does Color Really Help in Dense Stereo Matching?*. In: Proceedings of Conference on 3D Data Processing, Visualization and Transmission.

[Blumenthal, 2009] Blumenthal 2009. *Tectonic Shifts Altering the Terrain at Google Maps.* Posted October 2009 at http://searchengineland.com/tectonic-shifts-altering-the-terrain-at-google-maps-27783, last visited September 2010.

[Borgefors, 1988] Borgefors G. 1988. *Hierarchical chamfer matching: a parametric edge matching algorithm.* In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 6, pp. 849–865.

[Boykov et al., 2001] Boykov, Y., Veksler, O. and Zabih, R. 2001. *Fast approximate energy minimization via graph cuts*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1222 – 1239.

[Bresenham, 1965] Bresenham, J.E. 1965. *Algorithm for computer control of a digital plotter*. In: IBM Systems Journal, vol. 4, no. 1, pp. 25–30.

[Brenner, 1999] Brenner, C. 1999. *Interactive modeling tools for 3D building reconstruction*. In: D. Fritsch & R. Spiller, eds, Proceedings of Photogrammetric Week, Herbert Wichmann Verlag, Heidelberg, pp. 23–34.

[Brenner, 2000] Brenner, C. 2000. *Dreidimensionale Gebäuderekonstruktion aus digitalen Oberflächenmodellen und Grundrissen*. Dissertation. Deutsche Geodätische Kommission, Reihe C, Nr. 520.

[Brenner et al., 2001] Brenner, C., Haala, N. and Fritsch, D. 2001. *Towards fully automated 3D city model generation*. In: In Automatic Extraction of Man-Made Objects from Aerial and Space Images III.

[Brenner, 2005] Brenner, C. 2005. *Building reconstruction from images and laser scanning*. In: International Journal of Applied Earth Observation and Geoinformation, Theme Issue on "Data Quality in Earth Observation Techniques", vol. 6, no. 3-4, Elsevier, pp. 187-198.

[Bretterbauer et al., 2003] Bretterbauer K., Weber R. *A Primer of Geodesy for GIS-Users*. Geowissensch. Mitteilungen, Schriftenreihe der Studienrichtung Vermessung & Geoinformation, TU-Vienna, vol. 64, p. 50 S.

[Brown, 1976] Brown, D.C. 1976. *The bundle adjustment — progress and prospects*. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 21, no. 3.

[Brunn, 1998] Brunn, A. 1998. *Techniques for automatic building extraction*. Third Course in Digital Photogrammetry, Institute for Photogrammetry at Bonn University and Landesvermessungsamt Nordrhein-Westfalen, Bonn, Germany.

[Bullard et al., 2000] Bullard, S.H. and Straka, T.J. 2000. *Basic Concepts in Forest Valuation and Investment Analysis*; Clemson University, Department of Forestry, 2nd edition.

[Burger et al., 2008] Burger, W. and Burge, M.J. 2008. Comparing Images – Digital Image Processing, Publisher Springer, London, p. 449.

[Burochin et al., 2010] Burochin, J.-P., Vallet, B., Tournaire, O. and Paparoditis, N. 2010. *A Formulation For Unsupervised Hierarchical Segmentation Of Façade Images With Periodic Models*. In: Proceedings of PCV 2010 - ISPRS Technical Commission III Symposium on Photogrammetry, Computer Vision and Image Analysis, vol. 38, part 3A, Saint Mandé, France, pp. 227-232.

[Canny, 1986] Canny, J. 1986.*A Computational Approach To Edge Detection*, In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679–698.

[Cech et al., 2007] Čech, J. and Šára, R. 2007. *Windowpane detection based on maximum a posteriori labeling.*Technical Report TR-CMP-2007-10, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic, 14 p.

[Cech et al., 2009] Čech, J. and Šára, R. 2009. *Languages for Constrained Binary Segmentation Based on Maximum A Posteriori Probability Labeling*. In: International Journal of Imaging Systems and Technology, vol. 19, no. 2, pp. 69-79.

[Chan et al., 2004] Chan, T.F. and Esedoglu, S. 2004. *Aspects of total variation regularized L1 function approximation*. In: SIAM Journal on Applied Mathematics, vol. 65, no. 5, pp. 1817–1837.

[Chaudhuri et al., 2007] Chaudhuri, D. and Samal, A. 2007. *A simple method for fitting of bounding rectangle to closed regions*. In: Pattern Recognition, vol. 40, no. 7, pp. 1981-1989.

[Chiang et al., 2008] Chiang, Y. and Knoblock, C. 2008. *Automatic Extraction of Road Intersection Position, Connectivity, and Orientation from Raster Maps.* In: 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS 2008), pp 267-276.

[Collins, 1996] Collins, R. 1996. *A space-sweep approach to true multi-image matching*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 358–363.

[Comaniciu et al., 2002] Comaniciu, D. and Meer, P. 2002. *Mean shift: A robust approach toward feature space analysis*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603–619.

[Cormen et al., 1990] Cormen, T., Leiserson, C. and Rivest, R. 1990. *Introduction to Algorithms*.In: MIT Press.

[Corona, 1988] Unknown Authors, 1988. The Corona Story.National Reconnaissance Office. Link: http://www.nro.gov/foia/docs/foia-corona-story.pdf, p. 182.

[Cox et al., 1996] Cox, I.J., Hingorani, S.L., Rao, S.B. and Maggs B.M. 1996. *A maximum likelihood stereo algorithm*. In: CVIU, vol. 63, no. 3, pp. 542–567.

[Cramer et al., 2009] Cramer, M. and Haala, N. 2009. *DGPF project: Evaluation of digital photogrammetric aerial bases imaging systems - overview and results from the pilot centre*. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information

Sciences,ISPRS Workshop on High-Resolution Earth Imaging forGeospatial Information, vol. 38-1-4-7/W5, Leibniz-University Hannover, Germany, unpaginated.

[Curless et al., 1996] Curless, B. and Levoy, M. 1996. *A volumetric method for building complex models from range images*. In: SIGGRAPH, pp. 303–312.

[Cyclomedia, 2008] Cyclomedia 2008. http://www.cyclomedia.nl. Last accessed 14th Nov. 2009.

[Dan, 1996] Dan, H. 1996. *Rekonstruktion generischer Gebaeudemodelle aus Punktwolken und deren Abbildungskorrekturen in Orthobildern.*Doctoral and Habilitation Theses, ETHZuerich, Switzerland.

[Dang et al., 1993] Dang T., Jamet, O. and Maître, F. 1993. *Interprétation et Restitution Automatique de Bâtiments en Milieu Péri-Urbain*. Revue Française de Photogrammétrie et Télédétection, No. 131 (3), pp. 3-12.

[DGPF, 2009] DGPF 2009. *Photogrammetrie, Fernerkundung, Geoinformation*. Journal of the German Society for Photogrammetry, Remote Sensing and Geoinformation DGPF, H. Mayer (editor), Schweizerbart'sche Verlagsbuchhandlung, Stuttgart.

[Donoser et al., 2010] Donoser, M., Riemenschneider, H. and Bischof, H. 2010. *Linked edges as stable region boundaries*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1665–1672.

[Dornaika et al., 2008] Dornaika, F. and Bredif M. 2008. *An efficient approach to building superstructure reconstruction using digital elevation maps*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 37, (Part 3A), Beijing, China.

[Dorninger et al., 2008] Dorninger, P. and Pfeifer, N. 2008. *A Comprehensive Automated 3D Approach for Building Extraction, Reconstruction, and Regularization from Airborne Laser Scanning Point Clouds*. In: Sensors, vol. 8, no. 11, pp. 7323–7343.

[Douglas et al., 1973] Douglas D. and Peucker T. 1973. *Algorithms for the reduction of the number of points required to represent a digitized line or its caricature*. The Canadian Cartographer 10, pp. 112-122.

[Dulong et al., 2010] Dulong, C., Filip, D., Früh, C., Lafon, S., Lyon, R., et al., 2010. *Google Street View: Capturing the World at Street Level*. In: IEEE Computer, vol. 43, no. 6, pp. 32-38.

[Ebner et al., 1991] Ebner, H., Fritsch, D. and Heipke, C., 1991. *Digital Photogrammetric Systems*, Herbert Wichman Verlag GmbH, Karlsruhe, 344 pages.

[Ebner et al., 1999] Ebner, H., Eckstein, W., Heipke, C. and Mayer, H. 1999. *Automatic Extraction of GIS Objects from Digital Imagery*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 32, Part 3-2W5.

[Elaksher, 2002] Elaksher, A. 2002. *Building extraction from multiple images*, Ph.D. thesis, Purdue University.

[Facet, 2008] Facet Technology 2008. http://www.facet-tech.com/. Last accessed 14th of Nov. 2009.

[Fan, 2010] Fan, H. 2010. *Integration of time-dependent features within 3D city model*.Dissertation, Technical University Munich, Germany, 106p.

[Firschein et al., 1997] Firschein, O. and Strat, T.M. 1997. *Radius: Image Understanding for Imagery Intelligence*. Technical reports prepared for the DARPA Image Understanding Program, 503 pages.

[Fischler et al., 1981] Fischler, M.A. and Bolles, R.C. 1981. *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. In: ACM, vol. 24, no. 6, pp. 381-395.

[Fisher et al., 1998] Fischer, A., Kolbe, T.H., Lang, F., Cremers, A.B., Förstner, et al. 1998. *Extracting Buildings from Aerial Images Using Hierarchical Aggregation in 2D and 3D*. In: Computer Vision and Image Understanding, vol. 72, no.2.

[Forberg, 2004] Forberg, A. 2004. *Generalization of 3D Building Data Based on a Scale-Space Approach*. In: Proceedings of the 20th Congress of the ISPRS, vol. 35, Part B, Istanbul, Türkei.

[Förstner et al., 1995] Förstner, W. and Weidner, U. 1995. *Towards Automatic Building Reconstruction from High Resolution Digital Elevation Models.* In: ISPRS Journal of Photogrammetry and Remote Sensing; pp. 38-49.

[Förstner et al., 2010] Förstner, W. and Ying Yang, M. 2010. *Plane Detection in Point Cloud Data*.Technical Report-IGG-P-2010-01.

[Frahm et al., 2010] Frahm, J.M. et al. 2010. *Building Rome on a cloudless day*. In: Proceedings IEEE Computer Society European Conference on In Computer Vision, Hersonissos, Greece.

[Frueh et al., 2004a] Frueh, C. and Zakhor, A. 2004. *An automated method for large-scale, ground-based city model acquisition*. In: International Journal of Computer Vision, vol. 60, no. 1.

[Frueh et al., 2004b] Frueh, C., Sammon, R. and Zakhor, A. 2004. *Automated texture Mapping of 3D City Models with Oblique Aerial Imagery*. In: Proceedings of 2nd International Symposium on 3D Data Processing, Visualization and Transmission,Thessaloniki, Greece, pp. 396-403.

[Fua et al., 1987] Fua, P. and Hanson. A. J. 1987. *Using generic geometric models for intelligent shape extraction*, In: Proceedings of DARPA Image Understanding Workshop, Los Angles, USA.

[Furukawa et al., 2009] Furukawa, Y., Curless, B., Seitz, S. and Szeliski, R. 2009. *Manhattan-world stereo*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, USA.

[Fusiello et al., 2000] Fusiello, A., Trucco, E. and Verri, A. 2000. *A compact algorithm for rectification of stereo pairs*. In: Machine Vision and Applications, vol. 12, no. 1, pp. 16-22.

[GALLUP ET AL., 2007] Gallup, D., Frahm, J.M., Mordohai, P., Yang, Q. and Pollefeys, M. 2007. *Real-time Plane-sweeping Stereo with Multiple Sweeping Directions*.In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[GARTNER ET AL., 2008] Gartner G. and Rehrl, K. Eds. 2008. *Location Based Services and TeleCartography II. From Sensor Fusion to Context Models.* 5th International Conference on Location Based Services and TeleCartography, Salzburg, Austria, Lecture Notes in Geoinformation and Cartography; 2009, XXIX, 456 p.

[GEIBEL ET AL., 2000] Geibel, R. and Stilla, U. 2000. *Segmentation of laser-altimeter data for building reconstruction: Comparison of different procedures*. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 33, Part B3, pp. 326-334.

[GOODCHILD, 2008] Goodchild, M. 2008. *Assertion and authority: the science of user-generated geographic content*. In: Proceedings of the Colloquium for Andrew U. Frank's 60th Birthday. GeoInfo, vol. 39.Department of Geoinformation and Cartography, Vienna University of Technology.

[GORTE, 2002] Gorte, B. 2002. *Segmentation of TIN-structured surface models*. In: Joint International Symposium on GeospatialTheory, Processing and Applications, Proceedings CD, Ottawa, Canada, 5 p.

[GÖSELE ET AL., 2010] Gösele M., Ackermann, J. , Fuhrmann, S., Klowsky, R., Langguth, et al. 2010. *Scene Reconstruction from Community Photo Collections*, In: IEEE Computer vol. 43, no 6, pp. 48-53.

[GRABNER ET AL., 2007] Grabner, M., Nguyen, T., Gruber-Geymayer, B.C. and Bischof, H. 2007. *Boosting-based car detection from aerial images.* In: ISPRS Journal of Photogrammetry and Remote Sensing, vol. 63, no. 3, pp. 382-396.

[GRENZDÖRFFER ET AL., 2007] Grenzdörffer G., Guretzki, M. and Friedlander, I. 2007. *Photogrammetric image acquisition and image analysis of oblique imagery - a new challenge for the digital airborne system PFIFF*. In: ISPRS Hannover Workshop 2007 - High-Resolution Earth Imaging for Geospatial Information, Hannover, Germany, 6 p.

[GRÖGER ET AL., 2007] Gröger G., Kolbe, T. and Czerwinski, A. 2007. *Candidate OpenGIS CityGML Implementation Specification (City Geography Markup Language)*, In: OpenGIS Engineering Report. 168 p.

[GRUBER ET AL., 1995] Gruber, M., Pasko, M. and Leberl, F. 1995. *Geometric Versus Texture Detail in 3-D Models of Real World Buildings*, In: *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser Verlag, Basel, Switzerland, pp. 189-198.

[GRUBER, 1997] Gruber, M. 1997. *Ein System zur umfassenden Erstellung und Nutzung dreidimensionaler Stadtmodelle*, Dissertation, Graz University of Technology, 135 p.

[GRUBER, 2007] Gruber, M. 2007. *UltraCamX, the New Digital Aerial Camera System by Microsoft Photogrammetry*. In: D. Fritsch & R. Spiller, eds, Proceedings of Photogrammetric Week, Herbert Wichmann Verlag, Heidelberg, Germany.

[Gruber et al., 2008a] Gruber, M., Ponticelli, M., Bernögger, S. and Leberl, F. 2008. *UltracamX, the Large Format Digital Aerial Camera System by Vexcel Imaging / Microsoft*. In: Proceedings of the International Congress on Photogrammetry and Remote Sensing, Beijing, In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 37, pp. 665-670.

[Gruber et al., 2008b] Gruber, M. and Reitinger, B. 2008. *UltraCamX and a new way of photogrammetric processing*.In Proceedings of the *Annual Convention of the* American Society for Photogrammetry and Remote Sensing ASPRS, Portland, USA, unpaginated DVD.

[Gruber-Geymayer et al., 2006a] Gruber-Geymayer, B.C., Klaus, A. and Karner, K. 2006. *Data fusion for classification and object extraction;* In: CMRT05, Joint Workshop of ISPRS and DAGM; Vienna, Austria, pp.125-130.

[Gruber-Geymayer et al., 2006b] Gruber-Geymayer, B.C., Zebedin, L. and Karner, K. 2006. *From Pixels to Buildings;* VRVis Research Center for Virtual Reality and Visualization, Graz; 6 p.

[Grün et al., 1995] Grün, A., Kübler, O. and Agouris, P. eds. 1995. *Man-Made Structures from Aerial and Space Imagery,* Proceedings of First International Workshop held at Monte Veritaa, Ascona, Switzerland. Birkhäuser-Verlag, Basel. p. 250.

[Grün et al., 1997] Grün, A., Baltsavias. E.P. and Henricsson, O. (eds.) 1997. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, Proceedings of Second International Workshop held at Monte Veritaa, Ascona, Switzerland. Birkhäuser-Verlag, Basel.

[Grün et al., 1998] Grün, A. and Wang, X., 1998. *CC-Modeler, A topology generator for 3-D city models*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 32, Part 4 "GISBetween Visions and Applications", Stuttgart, Germany.

[Grün, 2009] Grün, A. 2009. *Photogrammetry – Revival of a Discipline*.GIS Development, pp. 30-31.

[Gu et al., 2009] Gu, C., Lim, J.J., Arbelaez, P. and Malik, J. 2009. *Recognition using regions*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1030-1037.

[Gunter et al., 1996] Gunter, J.E., Haney, H.L. Jr. 1996. Essentials of Forestry Investment Analysis; self published at University of Georgia and Virgina Tech; Second Edition.

[Gülch et al., 1999] Gülch, E., Müller, H. and Läbe T. 1999. *Integration of automatic processes into semi-automatic building extraction*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 32, Part 3-2W5 "Automatic Extraction of GIS Objects from Digital Imagery", Munich, Germany, pp. 177- 186.

[Gülch, 2000] Gülch, E. 2000. *Digital Systems for automated cartographic feature extraction*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 33, Part B2, pp. 241-255.

[HAALA ET AL., 1998] Haala, N., Brenner, C. and Anders, K.-H. 1998. *Generation of 3D City Models from Digital Surface Models and GIS*. In: ISPRS Workshop on 3D Reconstruction andModelling of Topographic Objects, Stuttgart, Germany, pp. 68-75.

[HAALA ET AL., 1999] Haala, N. and Brenner, C. 1999. *Extraction of buildings and trees in urban environments*, In: ISPRS Journal of Photogrammetry and Remote Sensing, vol. 54, no. 2-3, pp. 130–137.

[HAALA, 1999] Haala, N. 1999. *Combining Multiple Data Sources for Urban Data Acquisition*.In: D. Fritsch & R. Spiller, eds, Proceedings of Photogrammetric Week, Herbert Wichmann Verlag, Heidelberg, Germany.

[HAALA ET AL., 2009] Haala, N. and Wolff, K. 2009. *Digitale photogrammetrische Luftbildkamerasysteme – Evaluation der automatischen Generierung von Höhenmodellen*. In: DGPF, vol. 18, Eckhardt Seyfert, Publisher, pp. 23-32.

[HAALA ET AL., 2010] Haala, N. and Kada, M. 2010. *An update on automatic 3D building reconstruction*. In: ISPRS Journal of Photogrammetry and Remote Sensing, vol. 65, pp. 570–580.

[HAIDER ET AL., 2007] Haider, A., Seifert, C., Jindal, N., Paletta, L. and Paar, G. 2007. *Window Detection in Facades*. In: International Conference on Image Analysis and Processing, vol. 14, pp. 837–842.

[HAMMOUDI ET AL., 2011] Hammoudi, K., Dornaika, F. and Paparoditis, N. 2011. *Generating Virtual 3D Model of Urban Street Facades by Fusing Terrestrial Multi-source Data*. *In: Proc. IEEE International Conference on Intelligent Environments*, Nottingham, UK.

[HAN ET AL., 2005] Han, F. and Zhu, S.-C., 2005. *Bottom-up/top-down image parsing by attribute graph grammar*. In: Proceedings IEEE Computer Society International Conference on In Computer Vision, vol. 10, Beijing, China, pp. 1778–1785.

[HARLEY ET AL., 1987] Harley, J.B. and Woodward, D., 1987. *Eds. Cartography in Prehistoric, Ancient, and Medieval Europe and the Mediterranean*. vol. 1 of The History of Cartography. Chicago and London: University of Chicago Press.

[HARRIE, 1999] Harrie, L.E. 1999. *The Constraint Method for Solving Spatial Conflicts in Cartographic Generalisation*.In. Cartography and Geographic Information Systems.

[HARRIS ET AL., 1988] Harris C., Stephens M. 1988. *A combined corner and edge detector*. In: Proceedings of the Alvey Vision Conference. vol. 15, 50 p.

[HARTLEY ET AL., 2004] Hartley, R. I. and Zisserman, A. 2004. *Multiple View Geometry in Computer Vision*.Cambridge University Press.

[HAUGEARD ET AL., 2009] Haugeard, J.-E., Philipp-Foliguet, S. and Precioso, F. 2009. *Windows and facades retrieval using similarity on graph of contours*. In: Proc. IEEE International Conference on Image Processing, vol. 16, Cairo, Egypt. in electronic format only.

[HELMS, 1998] Helms, J.A. 1998. *The Dictionary of Forestry*. Society of American Foresters

[Henricsson, 1998] Henricsson, O. 1998. *The role of color attributes and similarity grouping in 3-D building reconstruction*. Computer Vision and Image Understanding, vol. 72, no. 2, pp. 163-184.

[Herman et al., 2010] Herman, M. and Kanade, T. 1986. *Incremental Reconstruction of 3D Scenes from Multiple, Complex Images*. In: Artificial Intelligence, vol. 30, no. 3, pp. 289-341.

[Hiep et al., 2009] Hiep, V.H., Keriven, R., Labatut, P. and Pons, J.P. 2009. *Towards high-resolution large-scale multi-view stereo*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[Hilton et al., 1996] Hilton, A., Stoddart, A.J., Illingworth, J. and Windeatt, T. 1996. *Reliable surface reconstruction from multiple range images*.In: Proceedings IEEE Computer Society European Conference on In Computer Vision, pp. 117–126.

[Hirschmüller, 2008] Hirschmüller, H. 2008. *Stereo Processing by Semiglobal Matching and Mutual Information*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 328-341.

[Hirschmüller et al., 2009] Hirschmüller, H. and Scharstein, D. 2009. *Evaluation of Stereo Matching Costs on Images with Radiometric Differences*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 9, pp. 1582-1599.

[Höhle, 2007] Höhle J. 2007. *Photogrammetric Measurements in Oblique Aerial Images*.In: Photogrammetrie, Fernerkundung, Geoinformation, vol. 1, pp. 7-14.

[Hohmann et al., 2009] Hohmann, B., Krispel, U., Havemann, S. and D. Fellner, D. 2009. *CityFit: High-quality urban reconstructions by fitting shape grammars to images and derived textured point clouds*. In: Proceedings of the 3rd ISPRS International Workshop 3D-ARCH "3D Virtual Reconstruction and Visualization of Complex Architectures" Trento, Italy.

[Hoover et al., 1996] Hoover, A., Fisher, R.B., Jean-Baptiste, G., Jiang, X., Flynn, P.J., et al. 1996. *An Experimental Comparison of Range Image Segmentation Algorithms*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[Hornung et al., 2006] Hornung, A. and Kobbelt, L. 2006. *Robust reconstruction of watertight 3D models from non-uniformly sampled point clouds without normal information*. In: Eurographics Symposium on Geometry Processing, pp. 41–50.

[Hsieh et al., 1996a] Hsieh, Y., 1996a. *SiteCity: A Semi-Automated Site Modelling System*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco 1996:

[Hsieh et al., 1996b] Hsieh, Y., 1996b. *Design and Evaluation of a Semi-Automated Site Modelling System*.In: Proceedings of DARPA Image Understanding Workshop, Palm Springs 1996.

[Huang et al., 1971] Huang, T.S., Schreiber W.F. and Tretiak O.J. 1971. *Image Processing*.In Proceedings of IEEE, vol. 59, no.11, pp. 1586-1609.

[Huertas et al., 1988] Huertas, A. and Nevatia, R. 1988. *Detecting buildings in aerial images*. In: Computer Vision, Graphics, and Image Process, vol.41, pp.131-152.

[Huertas et al., 1998] Huertas, A., Kim, Z. and Nevatia, R. 1998. *Use of IFSAR with Intensity Images for Automatic Building Modeling*.In: Proceedings of DARPA Image Understanding Workshop.

[Hui, 2009] Hui, J. 2009. *Optimization and Evaluation of the Microsoft Automatic Aerial Triangulation Software Package*. DiplomaThesis, Technical University of Stuttgart, Institute for Photogrammetry.

[Husch et al., 2002] Husch, B., Beers, T.W. and Kershaw J.A. Jr. 2002. *Forest Mensuration*. Wiley, 4 edition.

[ILMF, 2009] ILMF 2009. *International Laser Mapping Forum* 2009, New Orleans, organized by the American Society for Photogrammetry and Remote Sensing ASPRS, http://www.lidarmap.org/.Last accessed on 6th November.

[Irschara et al., 2007] Irschara, A., Zach, C. and Bischof, H. 2007. *Towards Wiki-based Dense City Modeling*.In: Proceedings IEEE Computer Society International Conference on In Computer Vision, Workshop on Virtual Representations and Modeling of Large-scale environments, pp. 1-8.

[Irschara, 2012] Irschara, A. 2012. *Scalable Scene Reconstruction and Image Based Localization*. Dissertation, Graz Technical University, p. 199.

[Irvin et al., 1989] Irvin.R. and McKeown D. 1989. *Methods for exploiting the relationship between buildings and their shadows in aerial imagery*. In: Proceedings IEEE-SMC, vol. 19, no. 6, pp. 1564 - 1575.

[Ishikawa, 2003] Ishikawa, H. 2003. *Exact optimization for markov random fields with convex priors*.In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 10, pp. 1333-1336.

[Jacobsen, 2010] Jacobsen, K. 2010. *Development of Digital Aerial Cameras*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 38, no. 1/W17, Istanbul, Turkey, 7 p.

[Jacobsen, 2011] Jacobsen, K. 2011. *Recent Developments of Digital Cameras and Space Imagery*. In: Ruzieka, J., Peskova, K. (Eds): Proceedings GIS Ostrava 2011, Technical University of Ostrava, 8 p.

[Jaynes et al., 1994] Jaynes, C., Stolle, F. and Collins, R. 1994. *Task Driven Perceptual Organization for Extraction of Rooftop Polygons*.In: Proceedings of DARPA Image Understanding Workshop, pp. 359-365.

[Jenks, 1981] Jenks, G.F. 1981. *Lines, computers and human frailties*. In: Annals of the Association of American Geographers, pp. 1-10.

[Jiang et al., 1996] Jiang, X.Y., Meier, U. and Bunke, H. 1996. *Fast Range Image Segmentation Using High-Level Segmentation Primitives*. In: Proceedings IEEE Computer Society 3rd Workshop on Applications of Computer Vision.

[Jin et al., 2005] Jin, H., Soatto, S. and Yezzi, A.J. 2005. *Multi-view stereo reconstruction of dense shape and complex appearance*. In: International Journal of Computer Vision, vol. 63, no. 3. pp. 175–189.

[Jochem et al., 2009] Jochem, A., Höfle, B., Rutzinger, M. and Pfeifer, N. 2009. *Automatic Roof Plane Detection and Analysis in Airborne Lidar Point Clouds for Solar Potential Assessment*. In: Sensors 9, pp. 5241-5262.

[Kada et al., 2009] Kada, M. and McKinley, L. 2009. *3D Building Reconstruction from Lidar Based on a Cell Decomposition Approach*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 38, Part3/W4, pp. 47–52.

[Karbo et al., 2009] Karbø, N., Schroth, R. 2009. *Oblique aerial photography: a status review*. In: D. Fritsch, eds, Proceedings of Photogrammetric Week, Herbert Wichmann Verlag, Heidelberg, Germany, pp. 119–125.

[Kassner et al., 2008] Kassner, R., Koppe, W., Schüttenberg, T. and Bareth, G. 2008. *Analysis of the Solar Potential of Roofs by using Official Lidar Data.*In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 37, Part B4. Vienna, Austria, pp. 399–403.

[Keriven et al., 1998] Keriven, R. and Faugeras, O. 1998. *Variational principles, surface evolution, PDE's, level set methods and the stereo problem*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 7, no. 3, pp. 336-344.

[Kienegger, 1992] Kienegger E. 1992. *Integration of Aerial Photographs with Geographic Information Systems.*Dissertation, Graz University of Technology.153 p.

[Kilian et al., 1996] Kilian, J., Haala, N. and Englich M. 1996. *Capture and evaluation of airborne laser scanner data* In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 31, no. B3, pp. 383-388.

[Klaus, 2007] Klaus A. 2007. *Object Reconstruction from Image Sequences*. Dissertation, Graz University of Technology, 147 p.

[Kluckner et al., 2009] Kluckner, S., Mauthner, T., Roth, P. and Bischof, H. 2009. *Semantic classification in aerial imagery by integrating appearance and height information*. In: Proceedings IEEE Computer Society Asian Conference on Computer Vision, unpaginated DVD.

[Kluckner, 2011] Kluckner, S. 2011. *Semantic Interpretation of Digital Aerial Images Utilizing Redundancy, Appearance and 3D Information.* Dissertation, Graz University of Technology, 180 p.

[Koivunen et al., 1991] Koivunen, V. and Pietikaeinen, M. 1991. *Combined edge- and region-based method for range image segmentation*. In: Proceedings of SPIE 1381, pp. 501.

[Kolbe et al., 2005] Kolbe, T.H., Gröger, G. and Plümer, L. 2005. *CityGML – Interoperable Access to 3D City Models*. In: Proceedings of the First International Symposium on Geo-Information for Disaster Management GI4DM, Delft, Netherlands.

[Kolbe, 2006] Kolbe, T.H. 2006. *CityGML – Background and Design*. Presentation of CityGML at OGC OWS-4, Reston, Virginia.(http://www.citygml.org/docs/Kolbe_CityGML_OWS-4_Kickoff_2006-06-14.pdf)..

[Kolbe, 2009] Kolbe, T.H. 2009. *Representing and Exchanging 3D City Models with CityGML*. In: Lee, J. and Zlatanova, S. (Eds.), 3D Geoinformation Sciences, Berlin, Germany, Springer.

[Kolbe et al., 2009] Kolbe, T.H., Nagel, C. and Stadler, A. 2009. *CityGML-OGC Standard for Photogrammetry?*In: D. Fritsch, eds, Proceedings of Photogrammetric Week, Herbert Wichmann Verlag, Heidelberg, Germany, pp. 265-277.

[Kolev et al., 2009] Kolev, K., Klodt, M., Brox, T. and Cremers, D. 2009. *Continuous Global Optimization in Multiview 3D Reconstruction*. In: International Journal of Computer Vision, pp. 80-96.

[Kolmogorov et al., 2004] Kolmogorov, V. and Zabi, R. 2004. *What Energy Functions Can Be Minimized via Graph Cuts?* In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 2.

[Korah et al., 2007] Korah, T. and Rasmussen, C. 2007. *2D Lattice Extraction from Structured Environments*. In: Proc. International Conference on Image Analyse and Recognition, vol. 2, Montreal, Canada, pp. 61–64.

[Kraus et al., 1998] Kraus, K. and Pfeiffer, N. 1998. *Determination of terrain models in wooded areas with airborne laser scanner data*, In: ISPRS Journal of Photogrammetry and Remote Sensing, vol. 53, pp. 193-203.

[Kraus, 2004] Kraus, K. 2004. *Photogrammetrie: Geometrische Informationen aus Photographien und Laserscanner Aufnahmen*. Band 1, De Gruyter Verlag.

[Kremer et al., 2007] Kremer J. and Hunter, G. 2007. *Performance of the Street- Mapper Mobile LIDAR Mapping System in „Real World" Projects*. In: D. Fritsch, eds, Proceedings of Photogrammetric Week, Herbert Wichmann Verlag, Heidelberg, Germany, pp. 215-225.

[Ladstätter et al., 2008] Ladstätter, R. and Gruber, M. 2008. *Geometric Aspects Concerning the Photogrammetric Workflow of the Digital Aerial Camera UltraCamX*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 37, no. 1, pp. 521–525.

[Lammi, 1999] Lammi, J. 1999. *A Method for Three-Dimensional Modelling of buildings from digital aerial imagery*.Dissertation. Acta Polytechnica Scandinavica, p. 117.

[Lafarge et al., 2006] Lafarge, F., Trontin, P., Descombes, X. and Pierrot-Deseilligny, M. 2006. *An automatic building extraction method: Application to 3D-city modelling*, INRIA research report No. 5925.

[Lafarge et al., 2010] Lafarge, F., Descombes, X., Zerubia, J. and Pierrot-Deseilligny, M. 2010. *Structural approach for building reconstruction from a single DSM*. In: IEEE Transactions on PatternAnalysis and Machine Intelligence, vol. 32, no. 1, pp. 135-147.

[Lafarge et al., 2011] Lafarge, F. and Mallet, C. 2011. *Building large urban environments from unstructured point data*.In: Proceedings IEEE Computer Society International Conference on In Computer Vision, Barcelona, Spain, November2011.

[Lang, 1969] Lang, T. 1969. *Rules for robot draughtsmen*. Geographical Magazine, vol. 22, pp. 50-51.

[Leberl et al., 2003a] Leberl, F. and Gruber, M. 2003. *Economical Large Format Aerial Digital Camera*. GIM International, The Worldwide Magazine forGeomatics.

[Leberl et al., 2003b] Leberl, F. and Gruber, M. 2003. *Flying the New Large Format Digital Aerial Camera UltraCamD*. In*:* D. Fritsch, eds, Proceedings of Photogrammetric Week, Herbert Wichmann Verlag, Heidelberg, Germany.

[Leberl et al., 2004] Leberl, F. and Thurgood, J. 2004. *The promise of Softcopy Photogrammetry Revisted*.

[Leberl et al., 2007] Leberl, F. 2007. *Die automatische Photogrammtrie für das Microsoft Virtual Earth*. In: ProceedingsInternationale Geodätische Woche Obergurgl. Chesi/Weinold(Hrsg.), Wichmann-Verlag, pp. 200-208.

[Leberl et al., 2009a] Leberl, F. and Gruber, M. 2009a. *Ortsbewusstsein im Internet – von 2-dimensionalen Navigationshilfen zur 3-dimensionalen Mixed Reality*.In: ProceedingsInternationale Geodätische Woche Obergurgl, Wichmann-Verlag, pp. 67-79.

[Leberl et al., 2009b] Leberl, F. and Gruber, M. 2009b. *3D-Models of the Human Habitat for the Internet*; In: Proceedings of Visigrapp-2009, Lisbon, Portugal; vol. IS, pp. 7-15.

[Leberl et al., 2010a] Leberl, F., Irschara, A., Pock, T., Meixner, P., Gruber, M., et. al. 2010. *Point Clouds: LiDAR versus 3D Vision*. In: Photogrammetric Engineering and Remote Sensing, vol. 76, no. 10, pp. 1123-1134.

[Leberl et al., 2010b] Leberl, F., Bischof, H., Pock, T., Irschara, A. and Kluckner, S. 2010. *Aerial Computer Vision for a 3D Virtual Habitat*. In: Proceedings IEEE Computer, vol. 43, no. 6, pp. 24-31.

[Leberl et al., 2010c] Leberl, F. 2010. *Time for Neo-Photogrammetry*. In: GIS Development, vol. 14, no. 2, pp. 22-24.

[Leberl et al., 2012] Leberl, F., Meixner, P., Wendel, A. and Irschara, A. 2012. *Urban 3D Modeling using Aerial Photography and 3D Vision*. In: Optical Engineering, In Press.

[Lee et al., 2004] Lee, S. C. and Nevatia, R. 2004. *Extraction and Integration of Window in a 3D Building Model from Ground View images*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, Minneapolis, USA, pp. 113–120.

[LEGO, 1967] Lego, K. 1967. Geschichte des österreichischen Grundkatasters. Bundesamt für Eich- und Vermessungswesen, Wien.

[LEICA, 2008] Leica 2008. Leica Scan Station 2: http://www.leicageosystems.com/hds/en/lgs_62189.htm. Last accessed 14th Nov. 2009.

[LEORDEANU ET AL., 2007] Leordeanu, M., Hebert, M. and Sukthankar, R. 2007. *Beyond local appearance: Category recognition from pairwise interactions of simple features*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[LI, 1996] Li, Z. 1996. *Transformation of Spatial Representation in Scale Dimension*. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 31, Part B3/III, pp. 453-458.

[LI ET AL., 1999] Li, J., Nevatia, R. and Nornoha, S. 1999. *User Assisted Modeling of Buildings from Aerial Images*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[LIN ET AL., 1994] Lin, C., Huertas, A. and Nevatia, R. 1994. *Detection of Buildings using Perceptual Grouping and Shadows*.In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 62-69.

[LIOW ET AL., 1994] Liow, Y. and Pavlidis., T. 1990. *Use of shadows for extracting building in aerial images*. In: *Computer Vision, Graphics and Image Processing*, vol. 49, pp. 242-277.

[LIU, 1998] Liu, J. 1998. *An ancient cadastre – Tongguan Tu*. In: Bulletin of Surveying and Mapping, vol. 7.

[LIU ET AL., 2010] Liu, M.Y., Tuzel, O., Veeraraghavan, A. and Chellappa, R. 2010. *Fast Directional Chamfer Matching*. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco.

[LORENZ ET AL., 2006] Lorenz, H. and Döllner, J. 2006. *Towards Automating the Generation of Facade Textures of Virtual City Models*. In: ISPRS Commission II, WG II/5 Workshop, Vienna ,Austria.

[LOWE ET AL., 1981] Lowe, D.G. and Binford, T.O. 1981. *The interpretation of Geometric Structure from image Motions.* In: Proceedings of DARPA Image Understanding Workshop, Washington DC, USA, pp. 47-60.

[LOWE ET AL., 2004] Lowe, D.G. 2004. *Distinctive image features from scale-invariant keypoints*. In: International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110.

[LUEBKE ET AL., 1995] Luebke, D. and Georges, C. 1995. Portals and Mirrors: *Simple, fast evaluation of potentially visible sets*. In: ACM Symposium on Interactive Graphics.

[LUEBKE ET AL., 2002] Luebke, D., Reddy, M. and Cohen, J.D. 2002. *Level of Detail for 3D Graphics*. In: Richardson, D. and Oosterom, P. (eds.), Advances in Spatial Data Handling 10th International, Symposium on Spatial Data Handling. Springer-Verlag, Berlin, pp. 137-152.

[Maas et al., 1999] Maas, H.-G.and Vosselman, G. 1999. *Two Algorithms for Extracting Building Models from Raw Laser Altimetry Data*.In: ISPRS Journal of Photogrammetry and Remote Sensing, vol. 54, no. 2-3, pp. 153-163.

[Marengoni et al., 1999] Marengoni, M., Jaynes, C., Hanson, A. and Riseman, E. 1999. *Ascender II, a Visual Framework for 3D Reconstruction*, In: First International Conference on Computer Vision Systems, Las Palmas, Spain.

[Martinec et al., 2007] Martinec D. and Pajdla, T. 2007. *Robust rotation and translation estimation in multi-view reconstruction*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8 p.*

[Matei et al., 2008] Matei, B., Sawhney, H., Samarasekera, S., Kim, J. and Kumar, R. 2008. Building segmentation for densely built urban regions using aerial lidar data. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, USA.

[Mayer, 1998] Mayer, H. 1998. *Model-Generalization of Building Outlines on Scale-Spaces and Scale-Space Events*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 32, no. 3/1, pp. 520-536.

[Mayer et al., 2005] Mayer, H. and Reznik, S. 2005. *Building Façade Interpretation from Image Sequences*. In: Proceedings of the ISPRS Workshop CMRT.

[McGlone et al., 2004] McGlone, C., Mikhail, E.M., Bethel, J. and Mullen, R. 2004. *Manual of Photogrammetry*, Fifth Edition, American Society for Photogrammetry and Remote Sensing ASPRS, 1151 p.

[McKeown et al., 1981] McKeown, D.M. Jr. and Kanade, T. 1981. *Database Support for Automated Photo Interpretation*.In: Proceedings of DARPA Image Understanding Workshop, Washington DC, USA, pp. 7-13.

[Merreli et al., 1981] Merrell, P., Akbarzadeh, A., Wang, L., Mordohai, P., Frahm, J.M., et al. 2007. *Real-time visibility-based fusion of depth maps*.In: Proceedings IEEE Computer Society International Conference on In Computer Vision.

[Metzger, 2006] Metzger, B. 2006. *Wertermittlung von Immobilien und Grundstücken*; Haufe Verlag, 2. Auflage.

[Meyer, 1989] Meyer, U. 1989. *Generalisierung der Siedlungsdarstellung in digitalen Situationsmodellen*. Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover p. 159.

[Milde et al., 2009] Milde, J. and Brenner, C. 2009. *Graph-based Modeling of Building Roofs*. In: Proceedings of AGILE Conference on GIScience, Hannover, Germany.

[Mohan et al., 1989] Mohan, R. and Nevatia, R. 1989. *Using perceptual organization to extract 3-D structures*.In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, pp. 1121-1139.

[Moons et al., 1998] Moons, T., Frère, D., Vandekerckhove, J. and Van Gool, L. 1998. *Automatic modelling and 3d reconstruction of urban house roofs from high resolution aerial imagery*.In: Proceedings IEEE Computer Society European Conference on In Computer Vision, pp. 410-425.

[Moore, 1965] Moore, G.E. 1965. *Cramming more components onto integrated circuits*. In: Electronics, vol. 38, no. 8.

[Müller et al., 2007] Müller, P., Zeng, G., Wonka, P. and Van Gool, L. 2007. *Image-based Procedural Modeling of Facades*. In: Proceedings of ACM SIGGRAPH/ ACM Transactions on Graphics, ACM Press, vol. 26, no. 3, article 85, 9 p.

[Nago et al., 1977] Nago, M., Natsuyamma, T. and Morr, H. 1977. *Structured Analysis of Complex Aerial Photographs*. In: Proceedings of the 6th International Joint Conference on Artificial Intelligence, pp. 610-616.

[Nago et al., 1978] Nago, M., Natsuyamna, T. and Ikeda, Y. 1978. Region Extraction and Shape Analysis of Aerial Photographs. In: Proceedings of the 14th International Conference on Pattern Recognition, pp. 620-628.

[Navratil et al., 2010] Navratil, G., Hafner, J. and Jilin, D. 2010. Accuracy Determination for the Austrian Digital Cadastral Map (DKM). In: Proceedings of the 4th Croatian Congress on Cadastre with International Participation (Medak, D., Pribicevic, B. & Delak, J., eds.), 15. - 17.2.2010, Zagreb, Croatia, pp. 171-181.

[Nevatia et al., 1973] Nevatia, R. and Binford, T.O. 1973. *Structured Descriptions of Complex Objects*. In: Proceedings of the 3rdInternational Joint Conference on Artificial Intelligence, Stanford University, USA.

[Nevatia et al., 1978] Nevatia, R. and Price, K. 1978. *Locating Structures in Aerial Images*. In: Proceedings of the 14th International Conference on Pattern Recognition, pp. 686-690.

[Nevatia et al., 1997a] Nevatia, R. and Huertas, A. 1997. *RADIUS Overview*. In: Proceedings of DARPA Image Understanding Workshop.

[Nevatia et al., 1997b] Nevatia, R. and Noronha, S. 1997. *Detection and Description of Buildings from Multiple Aerial Images*. In: Proceedings of DARPA Image Understanding Workshop, pp. 989-998.

[Nister, 2004] Nister, D. 2004. *An efficient solution to the five-point relative pose problem*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[Nister et al., 2006] Nister, D. and Stewénius, H. 2006. Scalable recognition with a vocabulary tree. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[Oda et al., 2004] Oda, K., Lu, W., Uchida, O. and Doihara, T. 2004. *Triangle-based Visibility Analysis and True Ortho Generation*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 35, Part B3, pp. 623-628.

[Okutomi et al., 1993] Okutomi, M. and Kanade, T. 1993. *A multiple-baseline stereo*. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 4, pp. 353-363.

[Ommer et al., 2010] Ommer, B. and Buhmann, J. 2010. *Learning the compositional nature of visual object categories for recognition*.In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32 no. 3, pp. 501-516.

[Opelt et al., 2006] Opelt, A., Pinz, A. and Zisserman, A. 2006. *Incremental learning of object detectors using a visual shape alphabet*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3-10.

[O'Reilly et al., 2009] O'Reilly T. and Batelle, J. 2009. *Web Squared: Web 2.0 Five Years On*.O'Reilly Media Inc. Available from www.web2summit.com.Last visited in October 2011.

[Paine et al., 2003] Paine, D.P. and Kiser, J. D. 2003. *Aerial Photography and Image Interpretation*. Wiley; 2 edition.

[Panday, 2011] Panday, U.S. 2011.*Fitting of parametric building models to oblique aerial images*.Master thesis, University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), Enschede, Netherlands.

[Paul, 2006] Paul, R. 2006. *Microsoft launches Virtual Earth 3D to try and take on Google Earth*. http://www.earthtimes.org/articles/show/10224.html, posted 7 Nov 06. Last visited September 2010.

[Petrie, 1998] Petrie, G. 1998. *Space Imagery for Topographic Mapping*. In: Geoinformatics, vol. 1, no.4, pp. 24-31.

[Petrie, 1999] Petrie, G. 1999. *Characteristics & Applications of High-Resolution Space Imagery*. In: Mapping Awareness, vol. 13, no. 10, pp. 33-37.

[Petrie, 2003] Petrie, G. 2003. *Airborne Digital Frame Cameras - The Technology is Really Improving*. GeoInformatics, Vol. 6, No. 7, p. 18-27.

[Petrie, 2008] Petrie, G. 2008. *Systematic Oblique Aerial Photography Using Multiple Digital Cameras*. In Proceedings of *VIII* International Scientific & Technical Conference: "From Imagery to Map: Digital Photogrammetric Technologies. Porec, Croatia.

[Petrie, 2009] Petrie, G. 2009. *Systematic Oblique Aerial Photography using Multiple Digital Frame Cameras*. In: Photogrammetric Engineering and Remote Sensing, vol. 75, no. 2, pp. 102-107.

[Peuquet, 2001] Peuquet, D.J. 2001. *Making Space for Time: Issues in Space-Time Data Representation*. In: GeoInformatica 5:1, pp. 11-32.

[Pluim et al., 2000] Pluim, J.P., Maintz, J.B. and Viergever, M.A., 2000. *Image Registration by Maximization of Combined Mutual Information and Gradient Information*. In: IEEE transactions on Medical Imaging, vol. 19, pp. 809-814.

[Pock et al., 2008] Pock, T., Schoenemann, T., Graber, G., Bischof, H. and Cremers, D. 2008. *A Convex Formulation of Continuous Multi-Label Problems.*In: Proceedings IEEE Computer Society European Conference on In Computer Vision, pp. 792-805.

[Pock et al., 2011] Pock, T., Zebedin, L. and Bischof, H. 2011. *TGV-fusion*. In: Rainbow of Computer Science. Springer Verlag. pp. 245-258.

[Polleyfeys et al., 2004] Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, et al. 2004. *Visual modeling with a hand-held camera*.In: International Journal of Computer Vision, vol. 59, no.3, pp. 207-232.

[Poullis et al., 2009] Poullis, C. and You, S. 2009. *Automatic reconstruction of cities from remote sensor data*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami, USA.

[Powitz, 1993] Powitz, B.M. 1993. *Zur Automatisierung der Kartographischen Generalisierung topographischer Daten in Geo-Informationssystemen.*Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen derUniversität Hannover, Germany, vol. 185.

[Prandi.,2008] Prandi, F. 2008. *Lidar and Pictometry Images Integrated Use for 3D Model Generation*. In: Proceedings of the 21th Congress of the ISPRS, Beijing, China.

[Price, 1981] Price, K.E. 1981. *Relaxation Matching Applied to Aerial Images*. In: Proceedings of DARPA Image Understanding Workshop, Washington DC, USA, pp. 26-31.

[Pu et al., 2007] Pu, S. and Vosselman, G. 2007. *Extracting windows from terrestrial laser scanning*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 36, part 3/W52.

[Quam et al., 1991] Quam, L. and Strat, T. 1991. *SRI image understanding research in cartographic feature extraction*.In: Ebner, H., Fritsch, D., Heipke, C. (Eds.), Digital Photogrammetric Systems, Wichmann Verlag, pp. 111-121.

[Ramer, 1972] Ramer, U. 1972. *An iterative procedure for the polygonal approximation of plane curves*. In: Computer Graphics and Image Processing, vol. 1, no. 3, pp. 244-256.

[Regnauld et al., 1999] Regnauld, N., Edwardes, A. and Barrault, M. 1999. *Strategies in Building Generalization: Modelling the Sequence, Constraining the Choice*. In: Progress in Automated Map Generalization.

[Reitinger et al., 2008] Reitinger B., Hoefler, M., Lengauer, A., Tomasi, R., Lamperter, M.,et. al 2008. *Dragonfly – Interactive Visualization of Huge Aerial Image Datasets*. In: Proceedings of the 21th Congress of the ISPRS in Beijing, China, pp. 491-494.

[Reitinger, 2008] Reitinger, B. 2008. *Interactive Visualization of Huge Aerial Image Datasets*; In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 37, Part B2 , pp. 491-494.

[Reumann et al., 1974] Reumann, K. and Witkam, A.P.M. 1974. *Optimizing curve segmentation in computer graphics*. In: Proceedings of the International Computing Symposium, pp. 467-472.

[Reznik et al., 2007] Reznik, S. and Mayer, H. 2007. *Implicit Shape Models, Model Selection, and Plane Sweeping for 3D Facade Interpretation*. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 36, Part 3A, Munich, Germany, p. 173.

[Ripperda et al., 2008] Ripperda, N., 2008. *Determination of Facade Attributes for Façade Reconstruction*. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 37, Part 3A, Beijing, China, pp. 285–290.

[Romberg et al., 1997] Romberg, J., Akram, W. and Gamiz, J. 1997. *Image Segmentation Using Region Growing*.Cited from internet source URL http://www.owlnet.rice.edu/elec539/projects97/WDEKnow/ index.

[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*. Institute for Photogrammetry and Remote Sensing at Vienna University of Technology, Vienna, Austria.

[Rottensteiner et al., 2002] Rottensteiner, F. and Briese, C. 2002. *A new method for building extraction in urban areas from high-resolution lidar data*. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 34, Part 3A, Graz, Austria, pp. 295-301.

[Rottensteiner et al., 2005] Rottensteiner, F., Trinder, J. and Clode, S. 2005. *Data acquisition for 3D city models from LiDAR*. In: Proceedings of the 25th IGARSS, Seoul, South Korea, pp. 521-524.

[Rottensteiner, 2007] Rottensteiner, F. 2007. *Consistent estimation of building parameters considering geometric regularities by soft constraints*. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 36, Part 3, Bonn, Germany, pp. 13-18.

[Rottensteiner, 2010] Rottensteiner, F. 2010. *Roof Plane Segmentation by Combining Multiple Images and Point Clouds*. In: Proceedings of PCV 2010 - ISPRS Technical Commission III Symposium on Photogrammetry, Computer Vision and Image Analysis, vol.38, Part 3A, Saint Mandé, France, pp. 245- 250.

[Roux et al., 1994] Roux, M. and McKeown, D.M. 1994. *Feature Matching for Building Extraction from Multiple Views*. In: proceedings. *IEEE Computer Vision and Pattern Recognition Conference*, 1994, 46-53.

[Russakoff et al., 2004] Russakoff, D., Tomasi, C., Rohlfing, T. and Maurer, G. 2004. *Image Similarity Using Mutual Information of Regions*. In: Proceedings IEEE Computer Society European Conference on In Computer Vision, Prague, Czech Republic, pp. 596–607.

[Saalfeld, 1999] Saalfeld, A. 1999.*Topologically consistent line simplification with the Douglas Peucker algorithm*. In: Cartography and Geographic Information Science, vol. 26, no. 1, pp. 7-18.

[Samadzadegan et al., 2005] Samadzadegan, F., Azizi, A., Hahn, M. and Lucas, C. 2005. *Automatic 3D object recognition and reconstruction based on neuro-fuzzy modelling*. In: ISPRS Journal of Photogrammetry and Remote Sensing, vol. 59, no. 5, pp. 255–277.

[Scharstein et al., 2003] Scharstein, D. and Szeliski, R. 2003. *A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms*.In: International Journal of Computer Vision, vol. 47, pp. 7-42.

[Schmid et al., 1997] Schmid, C. and Zisserman, A. 1997. *Automatic line matching across views*.In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, *Puerto Rico*, pages 666–671.

[Scholz et al., 2008] Scholz, S. and Gruber, M. 2008. *Radiometric Quality of UltraCam Images*.In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 37, Part B6a, Beijing, China, pp. 539-543.

[Scholz et al., 2009] Scholz, S. and Gruber, M. 2009. *Radiometric and Geometric Quality Aspects of the Large Format Aerial Camera UltraCamXp*. In: Proceedings of the ISPRS, Hannover Workshop 2009 on High-Resolution Earth Imaging for Geospatial Information, vol. 38-1-4-7/W5, pp. 143-147.

[Seitz et al., 2006] Seitz, S., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R. 2006. *A comparison and evaluation of multi-view stereo reconstruction algorithms*. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 519–526.

[Selfridge et al., 1981] Selfridge, P.G. and Sloan, K.R. Jr. 1981. *Reasoning About Images*. Application to Aerial Image Understanding; In: Proceedings of DARPA Image Understanding Workshop, Washington DC, USA, pp. 1-6.

[Sester, 2000] Sester, M. 2000. *Maßstabsabhängige Darstellung in digitalen räumlichen Datenbeständen*. Habilitationsschrift, Deutsche Geodätische Komission, Reihe C, Heft 544.

[Shannon, 1949] Shannon, C. E. 1949. *Communication in the presence of noise*. In: Proceedings of the Institute of Radio Engineers (IRE) / IEEE vol. 37, pp. 10–21.

[Simmons et al., 2007] Simmons, G. and Karbo, N. 2007. *Aerial Imagery from different Angles*. In: Professional Surveyor, vol. 27, no. 5, 4p.

[Smith et al., 1996] Smith, D.M., Larson, B.C., Kelty, M.J. and Ashton, P.M.S. 1996. *The Practice of Silviculture*. Applied Forest Ecology, Wiley, 9 edition

[Snavely et al., 2006] Snavely, N., Seitz, S.M. and Szeliski, R. 2006. *Photo tourism: Exploring photo collections in 3D*. In: Proceedings of ACM SIGGRAPH/ ACM Transactions on Graphics, ACM Press, vol. 25, no. 3, pp. 835–846.

[Snavely et al., 2008] Snavely, N., Seitz, S.M. and Szeliski, R. 2008. *Modeling the world from Internet photo collections*.In: International Journal of Computer Vision, vol. 80, no. 2, pp. 189-210.

[Snavely et al., 2010] Snavely, N., Simon, I., Goesele, M., Szeliski, R. and Seitz, S. 2010. *Scene reconstruction community photo collections*. In: Proceedings of the IEEE, vol. 98, no. 8, pp. 1370-1390.

[Staufenbiel, 1973] Staufenbiel, W. 1973. *Zur Automation der Generalisierung topographischer Karten mit besonderer Berücksichtigung großmaßstäbiger Gebäudedarstellungen*, Wissenschaftliche Arbeiten der FachrichtungVermessungswesen der Universität Hannover.

[Stilla, 1999] Stilla, U. 1999. *Reconstruction of building models from maps and laser altimeter data*. In: Proceedings of International Workshop for Integrated spatial databases: Digital images and GIS, Portland, USA.

[Stiny et al., 1972] Stiny, G. and Gips, J. 1972. *Shape Grammars and the Generative Specification of Painting and Sculpture*. In: Proceedings of IFIP Congress, vol. 71, Amsterdam, Netherlands, pp. 1460–1465

[Strassenburg-Kleciak et al., 2007] Strassenburg-Kleciak, M. 2007. *Photogrammetry and 3D Car Navigation*.In: D. Fritsch, eds, Proceedings of Photogrammetric Week, Herbert Wichmann Verlag, Heidelberg, Germany, pp. 309-314.

[Strecha et al., 2008] Strecha C., von Hansen, W, Van Gool, L., Fua, P. and Thoennessen, U. 2008. *On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery*.In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, USA, pp 1-8.

[Streetmapper] StreetMapper (undated) http://www.streetmapper.net/. Last accessed 14th Nov. 2009.

[Studholme et al., 1999] Studholme, C., Hill, D.L.G., Hawkes, D.J. 1999. An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. In: Pattern Recognition, vol. 32, no. 1, pp. 71–86.

[Stuehmer et al., 2010] Stuehmer, J., Gumhold, S. and Cremers, D. 2010. *Real-Time Dense Geometry from a Handheld Camera*, In: Proceedings of DAGM.

[Suesse et al., 2001] Suesse, H. and Voss, K. 2001. *A new efficient algorithm for fitting of rectangles and squares*. In: IEEE Computer Society International Conference on Image Processing.

[Sunday] Sunday (undated). Geometric algorithms. [last accessed on 12.11.2011].http://geometryalgorithms.com/Archive/algorithm0205/algorithm0205.htm

[Suveg et al., 2004] Suveg, I. and Vosselman, G., 2004. Reconstruction of 3D building models from aerial images and maps. In: ISPRS Journal of Photogrammetry and Remote Sensing, vol. 58, no. 3-4, pp. 202-224.

[Szeliski et al., 1999] Szeliski, R. and Golland, P. 1999. Stereo matching with transparency and matting. In: International Journal of Computer Vision, vol. 32, no. 1.

[TAILLANDIER ET AL., 2004] Taillandier, F. and Deriche, R., 2004. Automatic building reconstruction from aerial images: a generic bayesian framework. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 35, Part B3, pp. 343–348.

[TAINTER ET AL., 1996] Tainter, F.H. and Baker, F.A. 1996. *Principles of Forest Pathology*; Wiley.

[TARSHA-KURDI ET AL., 2007A] Tarsha-Kurdi, F., Landes, T. and Grussenmeyer, P. 2007a. *Hough transform and extended RANSAC algorithm for automatic detection of 3D building roof planes from Lidar data*, In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 36, Part 3 / W52.

[TARSHA-KURDI ET AL., 2007B] Tarsha-Kurdi, F., Landes, T., Grussenmeyer, P. 2007b. *Joint combination of point cloud and dsm for 3D building reconstruction using airborne laser scanner data*. In: Proceedings of 6th InternationalSymposium on Remote Sensing of Urban Areas, Paris, France.

[TARSHA-KURDI ET AL., 2007C] Tarsha-Kurdi, F., Landes, T., Grussenmeyer, P., Koehl, M. 2007c. *Model-driven and data-driven approaches using lidar data: analysis and comparison*. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 36, pp. 87-92.

[THIEMANN ET AL., 2004] Thiemann, F. and Sester, M. 2004. *Segmenation for Buildings for 3D Generalization*. In: ICA Workshop on Generalization and Multiple Representation, Leicester, UK.

[THIEMANN ET AL., 2006] Thiemann, F. and Sester, M. 2006. *3D-Symbolization Using Adaptive Templates*. In: Proceedings of ISPRS Technical Commission II Symposium 2006, Vienna, Austria, pp. 49-54.

[TOBLER, 1959] Tobler, W.R. 1959. *Automation and Cartography*. In: Geographical Review, vol. 49, pp. 526-534.

[TOBLER, 1964] Tobler, W.R. 1964. *An experiment in the computer generalization of map*.Technical report, Office of Naval Research, Geography Branch.

[TOLDO ET AL., 2008] Toldo, R. and Fusiello, A. 2008. *Robust Multiple Structures Estimation with J-linkage*. In: Proceedings IEEE Computer Society European Conference on In Computer Vision, Part 1, pp. 537-547.

[TOMLINSON ET AL., 1967]Tomlinson, R.F. 1967. *An introduction to the geo-information system of the Canada Land Inventory.*ARDA, Canada Land Inventory, Department of Forestry and Rural Development, Ottawa, Canada.

[TOSHEV ET AL., 2010] Toshev, A. Mordohai, P. and Taskar, B. 2010. *Detecting and parsing architecture at city scale from range data*. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA.

[TOTH, 2009] Toth, S.F.C. (eds.) 2009. *Topographic Laser Ranging and Scanning*.CRC Press, Taylor and Francis Group.590 p.

[Triggs et al., 1999] Triggs, B., McLauchlan, P.F., Hartley, R.I. and Fitzgibbon, A. 1999. *Bundle adjustment — a modern synthesis*. In: Proceedings of International Workshop on Vision Algorithms, pp. 298–372.

[Ullmann et al., 2007] Ullman, S. 2007. *Object recognition and segmentation by a fragment-based hierarchy*. In: Trends in Cognitive Sciences, vol. 11, no. 2, pp. 58 – 64.

[Ullrich et al., 2002] Ullrich, A., Studnicka, N. and Riegl, J. 2002. *Long-range highperformance time-of-flight based 3-d imaging sensors*, In: Proceedings of Conference on 3D Data Processing, Visualization and Transmission, pp. 852–855.

[Unger, 2012] Unger, M. 2012. *Convex Optimization for Image Segmentation*. Dissertation, Graz University of Technology.

[Van Kreveld et al., 2001] Van Kreveld, M. 2001. *Smooth Generalization for Continuous Zooming*. In: Proceeding of the ICA, Fourth Workshop on Progress in Automated Map Generalization, Peking, China.

[Venkateswar et al., 1990] Venkateswar.V. and Chellappa. R. 1990. *A framework for interpretation of aerial images*, In: Proceedings IEEE Computer Society International Conference on Pattern Recognition, Atlantic City, USA, pp. 204- 206.

[Visvalingam et al., 1993] Visvalingam, M. and Whyatt, J.D. 1993. *Line generalisation by repeated elimination of points*. In: Cartographic Journal, vol. 30, part 1, pp.46-51.

[Vosselman, 1999] Vosselman, G. 1999. *Building reconstruction using planar faces in very high density height data*. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 32, Part 3-2W5, Munich, Germany, pp. 87-92.

[Vosselman et al., 1999] Vosselman, G. and Veldhuis, H. 1999. *Mapping by Dragging and Fitting of Wire-Frame Models.* In: Photogrammetric Engineering and Remote Sensing, vol 65, no. 7, pp. 769-776.

[Wahl et al., 2007] Wahl, R., Klein, R. and Schnabel, R. 2007. *Efficient RANSAC for Point-Cloud Shape Detection*In: Computer Graphics Forum, vol. 26, no. 2, pp. 214-226.

[Wang et al., 2008] Wang, Y., Schultz, S. and Giuffrida, F. 2008. *Generation of Orthophotos Using Pictometry's Digital Images*. In: Proceedings of the *Annual Convention of the* American Society for Photogrammetry and Remote Sensing ASPRS, Portland, USA.

[Wang et al., 2011] Wang, R., Bach, J. and Ferrie, F. 2011. *Window Detection from Mobile LiDAR Data*, In Proceedings of IEEE Workshop on Applications of Computer Vision (WACV), Kona, Hawaii.

[Weibel et al., 1996] Weibel, R. 1996. *A Typology of Constraints to Line Simplification*. In: Proceedings of the 7th International Conference on Spatial Data Handling, pp. 533-546.

[Weiser.,1991] Weiser, M. 1991. *The Computer for the 21st Century.* In: Scientific American, Special Issue on Communications, Computers, and Networks, vol. 265, no. 3.

[WENDEL ET AL., 2009] Wendel, A. 2009. *Facade Segmentation from Streetside Image.*Master Thesis, Graz University of Technology, 72 p.

[WENDEL ET AL., 2010] Wendel, A., Donoser, M. and Bischof, H. 2010. *Unsupervised Facade Segmentation using Repetitive Patterns.* In: Proceedings of DAGM, LNCS 6376, Springer.

[WENDEL ET AL., 2011] Wendel, A., Irschara, A. and Bischof, H. 2011. Automatic Alignment of 3D reconstructions using a Digital Suface Model.In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Colorado Springs, USA, pp. 26-36.

[WENZEL ET AL., 2008] Wenzel, S., Drauschke, M. and Förstner, W. 2008. *Detection of repeated structures in facade images*. In: Pattern Recognition and Image Analysis, vol. 18, no. 3, pp. 406–411.

[WESSEL, P. ET AL. 1996] Wessel, P. and Smith, W.H.F., 1996. *A global self-consistent, hierarchical, and high-resolution shoreline database*. Journal of Geophysical Research, vol. 101, pp. 8741-8743

[WILLKOMM ET AL., 2009] Willkomm, P. 2009. *3D GDI - Automationsgestützte Erzeugung und Verteilung landesweiter Gebäudemodelle aus Laserdaten.* In: Proceedings of 14th Münchner Fortbildungsseminar GIS, Technische Universität München, unpaginated DVD.

[WONKA ET AL., 2003] Wonka, P., Wimmer, M., Sillion, F. and Ribasky, W. 2003. *Instant Architecture*. In: Proceedings of ACM SIGGRAPH/ ACM Transactions on Graphics, ACM Press, vol. 22, no. 3, pp. 669-677.

[WU ET AL., 2004] Wu, S.T., Adler, C.G. and Márquez, R.G. 2004. *The Douglas-peucker algorithm: sufficiency conditions for non-self-intersections*In: Journal of the Brazilian Computer Society, vol. 9, no.3.

[XIAO ET AL., 2010] Xiao, J., Gerke, M. and Vosselman, G. 2010. *Automatic detection of buildings with rectangular flat roofs from multi-view oblique imagery*. In: Proceedings of PCV 2010 - ISPRS Technical Commission III Symposium on Photogrammetry, Computer Vision and Image Analysis, vol. 38, Saint Mandé, France, pp. 251-256.

[XU ET AL., 1993] Xu, L. and Oja, E. 1993. *Randomized Hough Transform (RHT): Basic Mechanisms, Algorithms and Complexities*. In: Computer Vision, Graphics, and Image Processing : Image Understanding, vol. 57, no. 2, pp. 131–154.

[YOON ET AL., 2006] Yoon, K.J. and Kweon, I.S. 2006. *Adaptive Support-Weight Approach for Correspondence Search*.In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 4, pp. 650-656.

[YUAN ET AL. 2010] Yuan, J. and Boykov, Y. 2010. *TV-Based Multi-Label Image Segmentation with Label Cost Prior*. In: Procedings of the British Machine Vision Conference, number 1, pp. 101.1-101.12. British Machine Vision Association.

[ZACH, 2007] Zach, C. 2007. *High Performance Modeling from Multiple Views using Graphics Hardware*. Dissertation, Graz University of Technology, 153p.

[Zach et al., 2007a] Zach, C., Pock, T. and Bischof, H. 2007a. *A Duality Based Approach for Realtime TV-L1 Optical Flow*. In: Proceedings of DAGM.

[Zach et al., 2007b] Zach, C., Pock, T. and Bischof, H. 2007b. *A Globally Optimal Algorithm for Robust TV-L1 Range Image Integration*.In: Proceedings IEEE Computer Society International Conference on In Computer Vision.

[Zach, 2008] Zach, C. 2008. *Fast and High Quality Fusion of Depth Maps. International Symposium on 3D Data Processing*. In: Proceedings of Conference on 3D Data Processing, Visualization and Transmission.

[Zebedin et al., 2006] Zebedin, L., Klaus, A., Gruber-Geymayer, B. and Karner, K. 2006. *Towards 3D map generation from digital aerial images.* In: ISPRS Journal of Photogrammetry and Remote Sensing, vol. 60, no. 6, pp. 413-427.

[Zebedin et al., 2008] Zebedin, L., Bauer, J., Karner, K. and Bischof, H. 2008. *Fusion of Feature- and Area-Based Information for Urban Buildings Modeling from Aerial Imagery*. In: Proceedings IEEE Computer Society European Conference on In Computer Vision.

[Zebedin, 2010] Zebedin, L. 2010. *Fusion of Area Based and Feature Based Data Sets for Urban Site Reconstruction*. Dissertation, Graz University of Technology.

[Zhang et al., 2005] Zhang, W. 2005. *Automatic Digital Surface Model (DSM) Generation from Linear Array Images*. Dissertation, Swiss Federal Institute of Technology Zurich, 219 p.

[Zhang et al., 2006] Zhang, W. and Kosecká, J. 2006. *Nonparametric estimation of multiple structures with outliers*. In: Dynamic Vision, Volume 4358 of Lecture Notes in Computer Science, Springer, pp. 60–74.

[Zhao et al., 1997] Zhao, Z. and Saalfeld, A. 1997. *Linear-time sleeve-fitting polyline simplification algorithms*.In Proceedings of AutoCarto 13, pp. 214-223.

[Zhou et al., 2010] Zhou, Q. and Neumann, U. 2010. *2.5d dual contouring: A robust approach to creating building models from aerial lidar point clouds*. In: Proceedings IEEE Computer Society European Conference on In Computer Vision, Hersonissos, Greece.

[Zoller + Fröhlich, 2007] Zoller + Fröhlich 2007. www.zf-laser.com, Last accessed 14th Nov. 2009.

[Zuliani et al., 2005] Zuliani, M., Kenney, C.S. and Manjunath, B.S. 2005. *The multiRANSAC algorithm and its application to detect planar homographies*. In: IEEE Computer Society International Conference on Image Processing.