DEUTSCHE GEODÄTISCHE KOMMISSION
bei der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 526

Automatic Recognition and 3D Reconstruction of Buildings from Digital Imagery

Von der Fakultät für Bauingenieur- und Vermessungswesen
der Universität Stuttgart
zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Abhandlung

vorgelegt von

M.Sc. Babak Ameri Shahrabi
aus Tehran

München 2000

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

ISSN 0065-5325 | ISBN 3 7696 9565 8

Adresse der Deutschen Geodätischen Kommission:

Deutsche Geodätische Kommission

Marstallplatz 8 • D – 80539 München

Telefon (089) 23031113 • Telefax (089) 23031-283/-100

E-mail hornik@dgfi.badw.de • http://www.dgfi.badw.de/dgfi/DGK/dgk.html

Prüfungskommission

Hauptberichter: Prof. Dr.-Ing.habil. Dieter Fritsch, Universität Stuttgart

Prof. Dr.sc.techn. Toni Schenk, Ohio State University

Tag der Einreichung: 06.12.1999

Tag der mündlichen Prüfung: 28.04.2000

Der Druck dieser Publikation wurde durch das Institut für Photogrammetrie der Universität Stuttgart finanziert

---

© 2000 Deutsche Geodätische Kommission, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,
die Veröffentlichung oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie) zu vervielfältigen

ISSN 0065-5325 ISBN 3769695658

To Matin & Parsa

Zusammenfassung

Innerhalb der letzten Jahre haben sich die Bereiche Computer Vision und Photogrammetrie immer stärker angenähert. Computer Vision zielt vor allem auf die Entwicklung von Algortihmen zur automatischen Gewinnung von Information aus Bildern; in photgrammetrische Prozesse wird häufig noch versucht, die benötigte Information so genau wie möglich zu extrahieren. Die Verbindung und gegenseitige Befruchtung beider Disziplinen verspricht dabei deutlich bessere Resultate als dies nur einem allein Bereich möglich wäre. Die vorgestellte Arbeit setzt deshalb auf die Integration von Methoden aus beiden Bereichen. Das Ziel ist es dabei Gebäude aus komplexen Luftbildszenen zu extrahieren. Der Schwerpunkt liegt somit auf der Entwicklung neuer Konzepte und robuster Verfahren zu einem hierarchischen, datengetriebenen Proze zur Rekonstruktion allgemeiner polyedrischer Gebäudemodelle durch die Integration von Prozessen der Photogrammetrie und der Computer Vision. Der vorgeschlagene Ansatz basiert zunächst auf der pragmatischen Annahme, dass sich nahezu alle Dächer von Gebäuden durch eine Kombination ebener Flächen darstellen lassen. Aus diesem Grund kann ein sogenanntes polyedrisches Modell zur näherungsweisen Repräsentation auch komplexer Gebäude im Rahmen des Rekonstruktionsprozessen verwendet werden. Um zuverlässige und genaue Ergebnisse zu gewinnen wird in dem Verfahren eine starke Verknüpfung zwischen dem 2D Bild und dem 3D Objektraum eingesetzt. Zunächst wird eine Randwertbeschreibung einer ersten Gebäudehypothese in einem datengetriebenen Bottom-up Ansatz ausgehend von einfachen qualitativen geometrischen Primitiven im Bildraum hin zu komplexeren quantitativen Modellprimitiven im Objektraum generiert. Anschlieend erfolgt die Verifikation der ersten, groben Gebäudehypothese in einem modellgetriebenen Top-down Ansatz, um dadurch zu einer verfeinerten, zuverlässigeren und genaueren geometrischen Beschreibung zu gelangen. Konzeptuell kann das gesamte Spektrum der Arbeit in die drei grundlegenden Schritte Erkennung, Rekonstruktion und Verifikation der Gebäudehypothesen unterteilt werden. Obwohl diese Unterteilung keine scharf definierten Grenzen festlegt, liefert sie doch eine nützliche Rahmen zur Kategorisierung und Beschreibung der einzelnen Prozesse, die die wesentlichen Komponenten des autonomen Bildanalysesystems bilden. Die Erkennung beginnt mit einer groben Segmentierung eines Digitalen Höhenmodells (DHM) auf der Basis eines morphologischen Dilatationsverfahrens. Dadurch werden Regionen innerhalb der Luftbilder zu bestimmen, in denen mit groer Wahrscheinlichkeit einzelne Gebäude enthalten sind (Interessensgebiete). Anschlieend werden geometrische Eigenschaften der Flächenelemente genutzt, um ebene Bereiche innerhalb der Interessensgebiete zu detektieren. Die extrahierten Pixel dienen dann als Saatpunkte eines Bereichswachstumsverfahrens, das mit Hilfe einer Kleinsten-Quadrate-Schätzung ebene Bereiche selektiert. Dadurch wird die Bildoberfläche in Regionen unterteilt, die ebene Dachflächen repräsentieren. Die Rekonstruktion zielt auf eine modellorientierte Repräsentation der Gebäude. Dazu werden zunächst die zuvor segmentierten 2D Bildbereiche in den 3D Objektraum projiziert, um dreidimensional bestimmte Dachpolygone zu erhalten. Dies geschieht durch einen robustes Verfahren zur Parameterschätzung, das im Rahmen dieser Arbeit entwickelt wurde. Anschlieend werden die topologischen Beziehungen zwischen den 3D Dachpolygonen bestimmt. Die Nachbarschaftsbeziehungen werden basierend auf Voronoi Diagrammen berechnet und beschreiben die Nachbarschaftsbeziehungen zwischen den Grundelementen der Dachstruktur. Basierend auf den berechneten Topologien werden zunächst kompatible d.h. koplanare, benachbarte 3D Polygone zu gröeren ebenen 3D Dachpolygonen vereinigt. Darüber hinaus werden Symmetrien bezüglich benachbarter Polygone definiert und als Attribute für die weitere Prozessierung gespeichert. Die resultierenden 3D Polygone mit ihren topologischen Beziehungen und daraus abgeleiteten Attributen werden innerhalb des Programmpakets POLY-MODELER zu einem ersten groben Gebäudemodells kombiniert. Dabei wird eine Randwertbeschreibung der vorläufigen Gebäudehypothese durch die Verschneidung benachbarter Polygonflächen generiert. Abschlieend wird ein verfeinertes Gebäudemodell im Rahmen der Hypothesenverifikation in einem Top-Down Prozess generiert. Dabei dient die grobe Gebäudehypothese zur Erzeugung von Vertrauenesintervallen im Bild, die einen Suchraum für korrespondierende 2D Bildelemente definieren und so die Verfikation und Verbesserung der groben Gebäudehypothese ermöglichen. Die Modellparameter werden durch eine simultane Einpassung der geometrischen 3D Modellprimitive an die 2D Bildprimitive bestimmt, wobei gleichzeitig die geometrische und topologische Modellinformation als externe und/oder interne Zwangsbedingung einbezogen wird.

Abstract

In recent years, the fields of photogrammetry and computer vision have naturally grown towards each other. Computer vision seeks to develop algorithms for automatic extraction of information from imagery, while photogrammetric routines force algorithms to obtain required information as precisely as possible. The confluence of these two disciplines promises to produce results much greater than the contributions of either field alone. The present work is an attempt for integrating the contribution of both fields for the collection of buildings from complex scenes of aerial images. It is mainly concerned with introducing the new concepts and development of robust methods in a hierarchical framework, for a data-driven reconstruction of generic plane-face building objects through the integration of computer vision and digital photogrammetric techniques.

The proposed approach is based on the pragmatic assumption that building roofs are composed of generic plane-surfaces, so that a plane-face solid model, commonly called *polyhedral*, can approximate a complex building and is used to support the reconstruction process. In addition, there is a strong coupling between 2D image and 3D object space in order to achieve reliable and precise results. This is realized by a mutual design approach. A boundary representation (b-rep) of a coarse building hypothesis is generated in a bottom-up, data-driven approach from simple qualitative geometric primitives in image domain to more complex quantitative model primitives in object domain. Subsequently, the hypothesis model verification is performed in a top-down model-driven process to determine a reliable and accurate geometric description of the 3D structure elements of a reconstructed coarse building model.

Conceptually, the entire spectrum of this research work can be divided into three fundamental steps of *recognition, reconstruction*, and *hypothesis verification*. Although this subdivision has no definitive boundaries, it does provide a useful framework for categorizing and describing the various processes that are essential components of an autonomous image analysis system. The recognition part starts with a coarse segmentation of DSM in order to label areas (regions of interest)within aerial images, which have a high expectancy of representing individual buildings. This process is based on a morphological top-hat transformation. Furthermore, geometric characteristics of surfaces are used to extract flat-pixel surface type within detected regions of interest. The extracted pixels serve as the seed regions to a least squares planar fit region growing algorithm to partition the image surface into meaningful plane-roof regional primitives. To move to the more model-oriented representation of the buildings, which is carried out in the reconstruction part, the intermediate extracted 2D plane-roof regions are projected back into the object space, called 3D plane-roof polygons. This is performed using a synthesis robust parameter estimator technique developed in this thesis. In order to describe the interrelation between these 3D geometric primitives, the Polygons Adjacency Relationships (PAR) is computed. The adjacency relationships are defined based on Voronoi diagrams and describe the topological properties, in particular the neighborhood relationships between the basic elements of the roof structure. Based on the computed PAR the compatible adjacent 3D polygons are merged into the larger 3D plane-roof polygon and its symmetry with respect to their adjacent polygons is also defined and stored as its attributes for further processes. These 3D primitives along their adjacency relationships information and derived attributes are input to the POLY-MODELER, where they are geometrically and/or topologically combined to generate the coarse building model. The POLY-MODELER is a new model generator tool, which is originally developed in this study. It generates the boundary representation (b-rep) of a coarse hypothesis building model using the 3D intersection of adjacent polygons. Finally, the modified, highly accurate building model is generated in the hypothesis verification process in a top-down fashion. The reconstructed coarse building undergoes a refinement process based on FBMV (Feature Based Model Verification) concept. Treating the generated coarse building hypothesis as evidence leads to a set of confidence intervals in image space that can be used as the search space to find the corresponding 2D image primitives and performing a consistency verification of the reconstructed coarse model. An important component of the FBMV method is the ability to solve the model parameters by simultaneously fitting all the geometric primitives of the 3D model into the corresponding 2D image features while at the same time the geometrical and topological model information is imposed into the process as external and/or internal constraints.

Contents

Zusammenfassung **5**

Abstract **6**

1 Introduction **13**
1.1 Needs for 3D Model of Landscape . . . 14
1.1.1 Building Objects are Prominent . . . 15
1.2 Objectives of the Thesis . . . 17
1.2.1 General Framework . . . 17
1.2.2 Achievements . . . 18
1.3 Thesis Outline . . . 19

2 Building Reconstruction in Literature **21**
2.1 Introduction . . . 21
2.2 Theme in Building Reconstruction . . . 21
2.2.1 Primary Data Sources and Cues . . . 22
2.2.2 Supporting Object Model: Specific vs. Generic . . . 23
2.2.3 Methodology . . . 24
2.3 Semi-Automated Methods . . . 26
2.4 Automated Methods . . . 27

3 Robust Parameter Estimation **28**
3.1 Introduction . . . 28
3.2 Least Squares Principles . . . 28
3.3 Robust Parameter Estimation Methods in Computer Vision . . . 32
3.3.1 M-Estimator . . . 32
3.3.2 Random Sampling . . . 35
3.3.3 Clustering . . . 36
3.3.4 Case Deletion Diagnostics . . . 37
3.3.5 Minimum Description Length . . . 37
3.4 Synthesis Robust Estimators . . . 37

4 From Pixels to Geometric Primitives **41**
4.1 Introduction . . . 41
4.2 Why Region-based Segmentation . . . 41
4.3 Geometric Characteristics of Surfaces . . . 43
4.3.1 Mean and Gaussian Curvatures . . . 44
4.4 Regions of Interest . . . 46

4.5 Iterative Region-based Segmentation . . . 48
4.5.1 Image Noise Estimation . . . 49
4.5.2 Flat Pixels Type Labeling . . . 50
4.5.3 Edge Pixels Extraction . . . 51
4.5.4 Seed Region Extraction . . . 52
4.5.5 Region Growing . . . 53
4.5.5.1 Error Tolerance Thresholds . . . 54
4.5.5.2 Relaxation Labeling in Region Analysis . . . 54

5 Generic Polyhedral-Like Model Reconstruction 57
5.1 Introduction . . . 57
5.2 Primary Roof Elements in 3D Object Space . . . 57
5.3 Polygons Adjacency Relationships . . . 60
5.4 Merging Compatible Adjacent 3D Polygons . . . 64
5.5 POLY-MODELER: Generic Polyhedral-Like Model Generator . . . 66
5.5.1 Basic Notation . . . 67
5.5.2 Mathematical Concept and Methodology . . . 70
5.6 Experiments and Results . . . 74

6 Feature Based Model Verification 77
6.1 Introduction . . . 77
6.2 Motivation . . . 77
6.3 Feature Based Model Verification . . . 78
6.4 Mathematical Foundation . . . 80
6.4.1 Image Based Observations . . . 81
6.4.1.1 Linearity: A Local Internal Geometric Constraint . . . 81
6.4.1.2 Connectivity: A Global Internal Topological Constraint . . . 84
6.4.2 Image-Object Based Observations . . . 85
6.4.2.1 Collinearity: A Global External Geometric Constraint . . . 85
6.4.3 Object Based Observations . . . 87
6.4.3.1 Coplanarity: A Global External Geometric Constraint . . . 87
6.4.3.2 Conditional Constraints . . . 87
6.4.4 Combined Least Squares Adjustment . . . 89
6.5 Experiments and Result . . . 89
6.6 Quality Assessment . . . 93

7 Disscusion and Future Directions 95
7.1 Conclusion . . . 95
7.2 Directions for Further Research . . . 97

Bibliography 98

A Experimental Results 105

Acknowledgements 109

CurriculumVitae 110

List of Figures

1.1 Proposed setup for automatic recognition and 3D reconstruction of building objects 18

1.2 Three different building roof structures a) gable roof, b) hipped-gable roof, c) complex roof . . . 19

3.1 Linear regression: fit 1) six of the seven points are selected as inliers and the best fitted line are obtained by RANSAC (courtesy of Fischler and Bolles 1981), fit 2) least squares estimation provides an erroneous solution 30

3.2 Physical analogy that illustrates the sensivity of least squares methods to outliers (courtesy of Schunck 1990) 30

3.3 Hampel three-part M-estimator functions: a) Minimum function $\rho(x)$, b) Influence function $\psi(x)$, c) weight function $w(x)$ 34

3.4 Estimated 3D plane-roof polygon of a building roof overlaid on the 3D perspective view of the correponding DSM: a) 2D plane-roof region overlaid on corresponding roof structure, b) corresponding 3D plane-roof polygon back projected into the object space based on a standard LS estimation process, c) corresponding 3D plane-roof polygon back projected into the object space based on the RANSAC process, d) corresponding 3D plane-roof polygon back projected into the object space based on the synthesis robust estimation techniques 38

4.1 Object structure represented by its geometric primitives 42

4.2 Eight fundamental surface types defined by mean and Gaussian curvatures signs (Courtesy of Besl 1988) 45

4.3 3D perspective view of an image wrapped over corresponding DSM 46

4.4 Morphological opening of DSM: a) grey value based DSM image, b) 3D perspective view of DSM, c) 3D perspective view of computed DTM 47

4.5 Extracted regions of interest: a) the 3D perspective of the extracted region wrapped over coresponding DSM, b) extracted region overlaid on the corresponding aerial image. 48

4.6 Computed edge pixels binary image. 52

4.7 Extracted seed region overlaid on corresponding region of interest 53

4.8 Extracted 2D plane-roof regions overlaid on corresponding buildings roof: a) gable roof structure building, b) hipped-gable roof structure building, .c) complex roof structure building. 56

5.1 Estimated 3D plane-roof polygons in object space: a) extracted 2D plane roof regions of a gable roof structure, b) estimated 3D plane-roof polygons overlaid on corresponding gable roof structure building in object space, c) extracted 2D plane roof regions, and d) estimated 3D plane-roof polygons overlaid on corresponding hipped-gable roof structure, e) extracted 2D plane roof regions, and f) estimated 3D plane-roof polygons overlaid on corresponding complex roof structure building. 59

5.2 Correspondence between a Voronoi diagram and its dual, Delaunay triangulation network, a) Voronoi diagram, b) corresponding Delaunay triangulation network 60

5.3 Delaunay triangulation network generated based on the central gravity points of corresponding 2D plane-roof polygons 61

5.4 Chamfer 3-4 mask proposed by Borgefors (1986), a) symmetric Chamfer 3-4 mask used for parallel process, b) Chamfer 3-4 forward and , c) Chamfer 3-4 backward masks used for sequential process 62

5.5 Computation of Voronoi diagram based on distance transformation, a) the central point of each polygon primitive is considered as kernel point, b), and c) show the corresponding distance image and Voronoi diagram respectively, d) the polygonal primitives (medium size) are considered as initial kernel points, e), and f) illustrate the corresponding distance image and Voronoi diagram, the adjacency relationships between primitives are changed, g) the complete polygonal primitives are considered as initial kernel points, g), and h) illustrate the corresponding distance image and Voronoi diagram, the adjacency relationships between primitives are significantly changed. 63

5.6 Computation of polygon adjacency relationships (PAR), a) initial kernel polygons, b) distance image, the gray area in the middle indicates the zero distance, c) generated Voronoi diagram, d) computed adjacency graph 64

5.7 Merging adjacent 3D plane-roof polygons, a) and c) the 3D perspective view of the primary 3D plane-roof polygons of a gable and a complex roof structure building respectively, b) and d) the corresponding 3D plane-roof polygons after merging operation 66

5.8 Polygon parameterization. 68

5.9 3D intersection of adjacent polygons. 68

5.10 Stitching operation of adjacent polygons 69

5.11 Analysis of the point of intersection between three adjacent 3D-polys gives an indication of polygon convexity, a) 3D-poly p_1, is a convex 3D-polygons, b) 3D-poly p_1, is a concave 3D-polygons 70

5.12 Analysis of concave polygons 72

5.13 An example of the roof modeling based on the 3D intersection of the adjacent plane-faces 73

5.14 Determination of the feasible region (extension of the plane-surfaces of the roof model) by simultaneous solution of a set of n linear inequality conditions. 74

5.15 3D building roof modeling, a), d), and g) extracted 3D plane-roof polygons of a gable, hipped-gable and a complex roof structure before roof modeling respectively, b), e), and h) 3D perspective view of the corresponding reconstructed 3D coarse building roof models, c), f), and l) reconstructed coarse building roof models overlaid on corresponding building roof structures in 2D image space. 76

6.1 Perspective projection of a building into corresponding images 79

6.2 Uncertainty buffer of homologous model edges in corresponding images 82

6.3 Selected edge-pixels during the first iteration of the estimation process 82

6.4 Regression of a 2D image edge to the representative edge-pixels 83

6.5 Selected edge-pixels during the last iteration of the estimation process 84

6.6 Intersection of two adjacent edges 84

6.7 Two orthogonal adjacent edges 88

6.8 Reconstructed gable roof structure building: a) reconstructed coarse building model overlaid on aerial image, b) reconstructed fine building model overlaid on aerial image, c) perspective view of the reconstructed 3D coarse building model, d) perspective view of the 3D fine building model. 90

6.9 Reconstructed hipped-gable roof structure building: a) reconstructed coarse building model overlaid on aerial image, b) reconstructed fine building model overlaid on aerial image, c) perspective view of the reconstructed 3D coarse building model, d) perspective view of the 3D fine building model. 91

6.10 Reconstructed complex roof structure building: a) reconstructed coarse building model overlaid on aerial image, b) reconstructed fine building model overlaid on aerial image, c) perspective view of the reconstructed 3D coarse building model, d) perspective view of the 3D fine building model. 92

6.11 Top view of a single complex building 93

A.1 Reconstructed coarse buildings overlaid on the corresponding aerial image 106

A.2 Perspective view of 3D reconstructed coarse buildings 106

A.3 final reconstructed buildings overlaid on the corresponding aerial image 107

A.4 Perspective view of the final 3D reconstructed buildings 107
A.5 Perspective view of 3D reconstructed coarse buildings overlaid on the DSM 108
A.6 Perspective view of the final 3D reconstructed buildings overlaid on the DSM 108

List of Tables

3.1 The number m of subsets required to ensure $p \geq 95\%$ for given u and ε, where p is the probability that all the data points selected in one subset are non-outliers (courtesy of Torr and Murray 1993). 36

3.2 Comparison of the 3D reference plane with the corresponding 3D plane-roof polygons computed based on different estimation process. 39

4.1 Maximum possible number of combination between s randomly selected elements from a set of q distinct elements 42

4.2 Eight basic surface types defined by mean and Gaussian curvature signs (courtesy of Besl 1988) . 50

6.1 Verified coarse building model based on two corresponding aerial images 94

6.2 Verified coarse building model based on four corresponding aerial images 94

Chapter 1

Introduction

Man has always attempted to build machines that could make life somewhat easier or more pleasant. The speed of this process has been ever accelerating since the beginning of the industrial revolution, and technological development nowadays is moving faster than ever. The key issue in this progress is undoubtedly the computer, which has found many applications in our contemporary lifestyle. This is so because it is an enormously efficient machine for managing data of all kind. Large memory carriers allow storage of hundreds of gigabytes of data, fast processors can make millions of calculations with it and internet connections allow fast and low-cost transmission. Computers certainly outperform humans at these tasks. However, they require that these tasks be unambiguously defined. In other words, to perform a specific task by a computer, an algorithm must exist to tell it exactly which operation is to be carried out at which level, without any possible confusion. This is where computers differ from living beings. They cannot think, guess or take responsibility for their action. They surely lack intuition, i.e. the ability to decide which action to take in case of doubt. The latter properties which can be attributed to living beings and not to computers can be described by *intelligence*, which enables us to interact with our environment. Information is received from our environment by several receptors, e.g. the eye, the ear, etc., and is led to the brain. The brain analyses and processes this information by matching it with previous experiences and similar information stored in a vast array of quickly accessible knowledge (our memory), and finally decides what action to take to react to the received information. Therefore, a computer must have sensing capabilities in order to enable it to interact promptly with its environment. Among these capabilities, *vision* has long been recognized as the one with the highest potential to be built in into a computer environment because of the availability, for quite some time now, of high-quality visual sensors that can easily be hooked up to the computers (Faugeras 1996).

Computer vision has emerged over the years as the discipline to develop the theoretical and algorithmic basis by which useful information about the world can be extracted and analyzed from the observed image(s), in an automated manner. It is a collection of processes that, to a varying degree, model the functionality of the human visual and cognitive system in order to exploit the specific or generic knowledge of the imaged object or scene. The required information can be related to the recognition of a generic object, the three-dimensional description of an unknown object, the position and the orientation of the observed object, or the measurement of any spatial property of an object, such as the distance between two of its distinguished points. So far, current computer vision approaches are limited to highly restricted scenes and to particular application domains, e.g., industrial settings where illumination conditions, the type of objects and camera positions are rigidly inhibited. This is due to several reasons, firstly the embedding of an object in a scene and the imaging process itself may introduce many different kinds of noise and distributions. Objects may be partially occluded by other objects, the scene may have particularly high contrast, the sensor may be particularly noisy, and so forth. Thus, computer vision is confronted with the problem of processing noisy measurements. This is a very serious problem since this initial uncertainty must be tracked through all the subsequent processes that are built up within the system in order to achieve the final result. Secondly, recovery of three-dimensional (3D) information about the shape of objects is difficult. This is due to the fact that this information is usually lost in the imaging process, which creates a two-dimensional (2D) representation of the 3D world. This 2D image is related in a complex way to the structure of the real world through the physics of image formation and its geometry. Therefore, computer vision is faced with the inverse problem of recovering the lost dimension from the 2D images. Thirdly, an automated vision system must be able to determine the appropriate transition from the more image-oriented, qualitative representation of the object in the lower levels to the more abstract model-oriented, quantitative representation of the object at the higher levels. A major problem in precise definition of the nature of this mapping is the modeling aspect. The system must contain models of what we consider as objects. In fact, the creation of definitive models is difficult due to enormous variations in the geometric and functional descriptions of the objects of interest. Therefore, the system has to be built up based upon the more complex generic object model, which in turn increases the complexity of the problem in hand. In the past few years, researchers have attempted to employ computers to perform some tasks within a certain margin of error that are considered to require intelligence such as teaching the computer to speak, to recognize words in natural speech, or to play football, which have lead to promising results.

Complementary to computer vision that is more in the favor of automation for knowledge extraction, photogrammetry supports high quality knowledge acquisition and precise and reliable 3D description of an 2D object taking the power of geometry, in particular 3D geometry into account. This is because photogrammetry is inherently a three-dimensional measurement technique and therefore in principle is able to meet the requirements (Fritsch 1999). It is one of the fundamental technologies that is needed to combine and fuse information from more than one image, thus increasing the reliability of the information extraction process. This is achieved because of the presence of automated camera calibration methods and orientation techniques that have been the groundwork of photogrammetry for decades. However, the extraction of scene knowledge remained largely a manual or, at best, a semi-automated process and requires a shift from the conventional techniques of photogrammetry to ones that are more compatible with real time, and fully automated constraints that are emphasized in computer vision.

In recent years, the fields of photogrammetry and computer vision have naturally grown towards each other. Computer vision seeks to develop the algorithms for automatic extraction of information while the photogrammetric routines force the algorithms to obtain required information as precisely as possible. The confluence of these two disciplines promises to produce results much greater than the contributions of either field alone (Strat 1994). The present work is an attempt for integrating the contribution of both fields for acquisition of GIS objects from complex scene of aerial images. This is realized by devoting a great deal of attention to 3D geometry, as well as the problem of uncertain data. Even though geometry plays a crucial role in this process, this geometry has to be built from noisy measurements, which requires special attention to the field of statistics and the theory of estimation.

1.1 Needs for 3D Model of Landscape

Modeling and 3D description of real world objects collected through an imaging system has become a topic of increasing importance as they are essential for a variety of applications. Namely telecommunication for planning of wireless networks in cities (Siebe & Büning 1997, Leberl, Walcher, Wilson & Gruber 1999), urban environmental planning and design to support the decision making processes for development projects (Danahy 1999, Lange 1999), virtual tourist information systems to support the on-line positioning, access and queries on the information of the site of interest (Volz & Klinec 1999), defense and military organization to support the training operation in virtual environment, architectural design for the realistic visualization of the drawing, environment and resource management, monitoring, and control for disaster preparedness, simulation of air pollution and noise distribution, to mention only a few. This broad range of applications and activities poses a number of issues and open questions that have to be discussed. It should be emphasized as well that 3D reconstruction and the representation of the geometry and shape of the world objects are important issues, their semantic information, administration and maintenance also need special attention. Although some efforts have been reported to clarify the common interests between the producers and users of this type of data (Fuchs, Gülch & Förstner 1998), however, more detailed study and strong research need to be tackled in the following themes:

1. Type of the objects e.g., buildings, roads, trees, etc., their geometric specification such as level of details and accuracy have to be defined for different type of applications. Although in practice, these prerequisites can largely mimic the conventions of traditional 2D maps and GIS, i.e., geometric accuracy requirements, there are new aspects that have to be taken into account. For example, the level of details in a traditional 2D map is limited to the exterior of the object of interest, while working in a 3D space allows reconstructing the interior of the object as well, therefore imposing novel extension to the acquisition of spatial data.

2. The geometrical modeling and visualization are the milestones of the field of virtual world model. Technical issues in structuring the 3D data e.g., vector or raster data structure, topics in geometric modeling such as boundary (e.g., b-rep), or volumetric (e.g., CSG) representation, which have a strong role in the domain of computer graphics have to be discussed and to a large extent standardized. The shortage, benefits, and efficiency of different solutions from algorithmic, methodological, and logistical points of view should be elaborated in the framework of a true 3D world model in the local (block, district) and global (city, regional) geographical extensions. For example, the 3D real-time representation of the cities (3D Urban GIS), where the user can visit the places, the streets, and the interior of the buildings as a virtual tourist, may require extensive visualization equipment (Gruber 1999). But its realization certainly depends on the geometric shape description and surface representation of the city objects and has to be supported with high quality texture mapped photo-realistic visualization.

3. Management, maintenance, fast interaction, data exchange policy and interoperability of the extraordinary large quantities of 3D data has to be investigated. If a system is designed to handle such data, including geometry, aerial and terrestrial photo-texture, and additional semantic information, as well as different viewing capabilities, it has to be aware of manipulating the hundreds of gigabytes of data. This introduces the new and interesting research challenges aiming to view the 3D data, zoom it, modify it, and query it with the objectives in mind to solve the problem based on a smart data organization (Kofler, Rehatschek & Gruber 1996), instead of utilizing special hardware configuration, in order to prevent the mobility characteristics of such a system, and its accessibility via the standard home computers by placing the reconstructed virtual models in the WWW home pages.

4. An important issue that has been neglected so far is the primary source of data. It is still an open question of how, when, where and why the different data sources fuse. In fact, a great variety of techniques and methods have already been reported by researchers (Haala & Anders 1996, Haala & Anders 1997), to convert various data sources such as satellite, aerial, or terrestrial images taken by optical (i.e. color or BW intensity-based images), laser (i.e., pulse or continuous wave range data), or microwave (i.e., Interferometric SAR) imaging sensors, existing 2D GIS and DTM, into the useful entries for generating the 3D spatial information system. However, the potentials, efficiency, and applicability, as well as the impact of any individual source material in this process need to be further evaluated. Moreover, the important issues of time, cost and availability of such data should be taken into account.

5. As a matter of fact, there are still other important issues to be cited and discussed, e.g., the mechanisms of data revision, but it would overload this introductory reading.

The discussion above reveals a strong motivation and evokes a challenging research effort for integration and interaction between different disciplines (photogrammetry, computer vision, computer graphics, database design and administration, as well as computer networking), which are required for shaping and developing a true 3D spatial information system. In fact, a possible way to decrease the problem complexity associated with such a system is to restrict the problem statement to a certain application area. For this reason, this study is only restricted to the domain of recognition and 3D reconstruction of building objects from stereo aerial images. However, the individual components and the complete framework as a whole are designed –whenever possible– with special attention and a careful study of the above mentioned requirements. For example, the process of reconstruction is based on a generic polyhedral-like object model, thus it would be possible to integrate acquisition of other world objects if they can be approximated by a polyhedral object model. The hypothesis verification process allows refinement of the final object model with a dynamic range of geometrical accuracy simply by tuning the thresholding parameters. In addition, the system conceptually is capable to alarm the cases that a visual control or a manual modification of the final result is required (traffic light concept (Förstner 1996)). Moreover, the geometric modeling process is based on a b-rep model, thus the geometric primitives along their topological information is kept which enables an easy transformation of the reconstructed objects into any specific format required by the end user.

1.1.1 Building Objects are Prominent

Building objects are recognized to be the most prominent objects in a 3D Urban Information System (UIS). *'Virtual reality and three-dimensional visualization are on the verge of changing the practice of urban environmental planning and design. Instead of presenting citizens with abstract maps and descriptive text to explain, analyze and debate design ideas and urban processes, planners will be able to show people explicit photo-textured information of what their city will look like after a proposed change ... Photo-textured 3D models are easy for people to understand quickly. They can recognize specific elements and orient their view in terms of spatial position and scale. Unless people have had a lot of experience reading plans, the traditional products of planning and GIS can be undecipherable or confusing to non-experts. This can leave people with the wrong impression of a design's positive and negative aspects'* (Danahy 1999, pp. 351–352). This quote is confirmed by the result of survey of the European Organization for Experimental Photogrammetric Research (OEEPE), on 3D city model. 95 % of the participants has reported that buildings are the most interesting and important objects in a 3D UIS (Fuchs et al. 1998). Consequently, a large number of research projects and efforts have been invested in the field of recognition, 3D reconstruction, and representation of building objects over the last few years (Collins, Hanson, Riseman & Schultz 1995, Faugeras, Laveau & Robert 1995, Förstner 1995, Gruber, Pasko & Leberl 1995, Haala 1995, Kim & Mueller 1995, Lin, Huertas & Nevatia 1995, Weidner & Förstner 1995, Axelsson 1996, Bignone, Henricson, Fua & Stricker 1996, Henricson, Bignone, Willhuhn &

Ade 1996, Weidner 1996, Englert 1997, Gruen & Dan 1997, Haala & Anders 1997, Hendrickx, Vandekerckhove, Frere, Moons & Gool 1997, Jaynes, Hanson & Riseman 1997, Kulschewski & Koch 1997, Nevatia, Lin & Huertas 1997, Stilla, Geibel & Jurkiewicz 1997, Fritsch & Ameri 1998, Brenner & Haala 1998*b*, Fischer, Kolbe, Lang, Cremers, Förstner, Plümer & Steinhage 1998, Gülch, Müller, Läbe & Ragia 1998, Haala, Brenner & Anders 1998, Ameri & Fritsch 1999, Baillard, Schmid, Zisserman & Fitzgibbon 1999, Brenner 1999, Fischer, Kolbe & Lang 1999, Förstner 1999, Gruen 1999, Haala 1999, Kulschewski & Koch 1999, Maas 1999, Nevatia, Huertas & Kim 1999, Vosselman 1999, Ameri & Fritsch 2000).

The acquisition and 3D representation of building objects includes not only the detection of the buildings in the scene depicted by one or more images, but also the production of a scene description. This is a complex task consisting of different processes such as recognition, feature extraction, feature's attributes computation, grouping, structuring and geometric modeling, hypothesis generation, as well as hypothesis verification, which are assembled in a cautious manner. As a matter of fact, integration of all these processes in order to derive the required 3D information from a complex scene image(s), in a traditional way, human-based image interpretation system, would be a costly and labour intensive operation (Duperet, Eidenbenz & Holland 1997). The laborious process of hand digitizing and interactively crafting each geometric model of buildings is too much time consuming. Therefore, there is an increasing demand towards fully machine-based image interpretation systems. This is a difficult task for several reasons:

1. The enormous variations in the structure and shape of the buildings prohibits using the specific object models or a constrained model to support the scene interpolation. Even imposing too tight constraints on the geometric regularity of the building structure, although it is an important component in architectural design of buildings, prevents the detection of many structures that do not satisfy them perfectly. This leads to selection of a generic object model, i.e., plane-face solid object model, in the expenses of increasing the complexity of the problem at hand.

2. Occlusion of buildings or building-parts by themselves or with neighboring adjacent objects such as buildings, trees, or cars cause that recognition process fails to provide complete or at best sufficient information for hypothesizing the major structure of the buildings. In fact, this problem can be partially overcome e.g., using the multiple images taken from different view points, again in the expenses of increasing complexity of the task and computational burden.

3. The effect of shadows, noise, low contrast, or small structures on the roof structure or presence of other objects leads to extraction of additional or spurious data which are not relevant and cause ambiguities or confusions in higher level processes of reconstruction and geometric modeling. This type of problems can also be partially overcome utilizing other cues, e.g., 3D geometry, or to constrain the problem, e.g., minimum size threshold, in the expenses of disregarding some of the relevant data.

4. Recovery of 3D information about the shape of the buildings is difficult. This is due to the fact that one geometric dimension is lost in the imaging process. In other words, the 2D spatial sampling process carried out by imaging sensor distorts the shape of the buildings in a non-reversible way. The homologous 2D primitives of building structure are related to the structure of the real world buildings through a complex relation based on the physics of image formation and its geometry. Therefore, the reconstruction process is faced with the inverse problem of recovering the lost dimension from the 2D images. However, using stereo image techniques, this relation can be established with a certain degree of precision using the well-known theory of perspective geometry.

5. The detection of individual buildings in downtown areas, where the individual buildings are attached and form the block of buildings, is an important issue in the particular application of automatic building reconstruction. In fact, the human operator discriminates two adjacent but distinct buildings simply based on the enclosure of the building contours. The current techniques for the detection of individual building such as Mathematical Morphologic Operations (see section 4.4), or image classification using height information as a supporting source of data (Walter 1999), at best are only capable of detecting the disjoint building objects. That means the fully automated reconstruction of buildings in densly built-up areas is only feasible if the individual buildings are signalized by the interaction of a human operator, or alternatively for the time being, utilizing the existing 2D GIS information such as ground plan of the individual buildings (Haala & Brenner 1999, Brenner 1999).

In spite of current limitations mentioned above, techniques used in photogrammetry and computer vision are now sufficiently developed and have resulted in sophisticated systems and led to promising result for acquisition and reconstruction of GIS objects, in particular buildings (see appendix A).

1.2 Objectives of the Thesis

This thesis addresses the problem of automatic detection and 3D reconstruction of buildings using aerial images. It is mainly concerned with introducing the new concepts and development of robust methods in a hierarchical framework, for a data-driven reconstruction of generic plane-face building objects through the integration of computer vision and digital photogrammetric techniques. The term *data-driven* is used to indicate that the process of recognition and reconstruction is performed without a priori knowledge about the building type or its structure, and the term *generic* is used to emphasize the fact that this type of reconstruction is not based on specific, user-defined building models, but rather on coarse, complex building models. It is designed to manage buildings of different shapes and complexities (exceptions are buildings with curve-like roof structure, such as dome roof). Thus, most of the geometrical regularity constraints imposed in the low- and intermediate-level reconstruction phases such as orthogonality, or parallelism, which are required in the specific model-based methods are not appropriate here. The proposed approach is only based on the pragmatic assumption that building roofs are composed of generic plane-surfaces, so that a plane-face solid model, commonly called *polyhedral* can approximate a complex building and is used to support the reconstruction process. In addition, there is a strong coupling between 2D image and 3D object space in order to achieve reliable and precise results. This is realized by a mutual design approach. A boundary representation (b-rep), of a coarse building hypothesis is generated in a bottom-up, data-driven approach from simple qualitative geometric primitives in image domain to more complex qualitative model primitives in object domain. Subsequently, the hypothesis model verification is performed in a top-down model-driven approach by back projecting the constructed model into the corresponding aerial images. The verification process is performed by simultaneously fitting the reconstructed model primitives into the homologous 2D features in images taken from different viewpoints while at the same time the geometrical and topological model information is imposed into the process as external and/or internal constraints.

The next section presents an overview of the whole framework. It gives a summary on the interrelated processing flow and concepts of our method for solving the task and the major contributions and achievement of this study. The outlines and structure of the thesis conclude this chapter.

1.2.1 General Framework

Figure (1.1), schematically represents the workflow of the subsequent processes and the interrelation between the major components of the automated reconstruction method proposed in this thesis. The proposed components form a general framework, in which in each step different and more complex types of information are exploited. Conceptually, the entire spectrum of our work can be divided into three fundamental steps of *recognition, reconstruction*, and *hypothesis verification*. Although this subdivision has no definitive boundaries, it does provide a useful framework for categorizing and describing the various processes that are essential components of an autonomous image analysis system.

The recognition part starts with a coarse segmentation of DSM in order to label areas within aerial images, which have a high expectancy of representing individual buildings. This process is based on a *morphological top-hat transformation,weidner:95*. Furthermore, geometric characteristics of surfaces, the *mean* and *Gaussian curvatures* are used to extract flat-pixel surface type (Besl & Jain 1988). The extracted 4-connected flat-pixels serve as the seed regions to a *least squares planar fit region growing algorithm* to partition the image surface into meaningful primitive plane-roof regions (Fritsch & Ameri 1998). To move to the more model-oriented representation of the buildings, which is carried out in the reconstruction part. The intermediate extracted 2D plane-roof regions are projected back into the object space, called 3D plane-roof polygons. The result is based on a *robust parameter estimator* developed within this research study. Estimating the initial parameters of the surface normal vector based on random sampling type estimators, Random Sample Consensus, *RANSAC* (Fischler & Bolles 1981), or alternatively Least Median Squares, *LMS* (Rousseeuw & Leroy 1987), we proceed with an iteratively re-weight M-estimator (Huber 1981, Hampel, Ronchetti, Rousseeuw & Stahel 1986). In order to describe the interrelation between these 3D geometric primitives, the *Polygon Adjacency Relationship (PAR)* is computed. The adjacency relationships are defined based on a Voronoi diagram (dual of Delaunay triangulation). Polygons are considered adjacent only if their Voronoi regions touch. The Voronoi region computation based on distance transformation in raster domain (Borgefors 1986), has extended in such a way that shape and boundary of the polygons are also taken into account. In this manner, topological information such as 'contained-in' relationships are computed more efficiently. Based on the PAR, the compatible adjacent 3D polygons are *merged* into the larger 3D plane-roof polygon. These 3D elements along their adjacency relationships information are input to the POLY-MODELER, where they are geometrically and/or topologically combined to

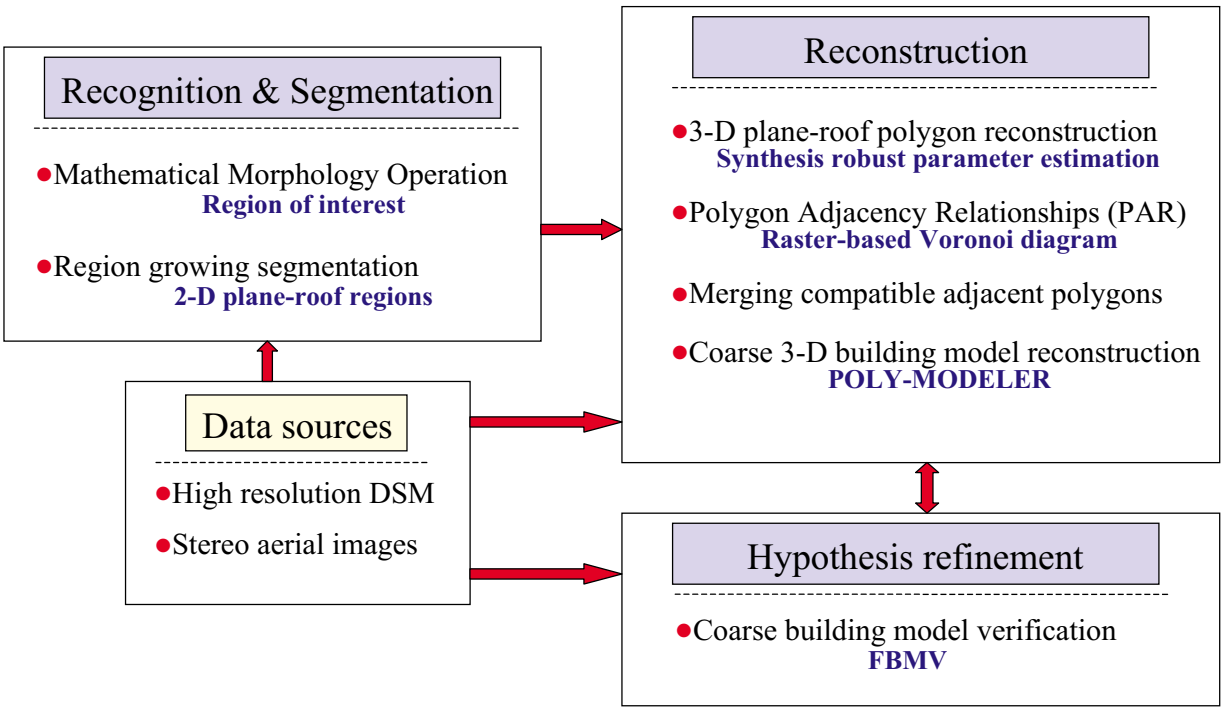


Figure 1.1: Proposed setup for automatic recognition and 3D reconstruction of building objects

generate the generic coarse polyhedral-like building model. Note that in this study the provided object model is generated in an image-driven, geometrically constrained process based on the intersection of adjacent 3D plane-roof polygons. This is in contrast to the general approach that a specific user-defined model is introduced into the system. Although, the specific models are attractive because of their ability to capture common symmetries and represent certain shapes with few parameters and most importantly are simple to work with, obviously they are inadequate for representing real-world objects that do not exhibit the set of regularities incorporated into the primitives. Finally, the modified, highly accurate building model is generated in the hypothesis verification process in a top-down fashion. The reconstructed coarse building undergoes a refinement process based on FBMV (Feature Based Model Verification) concept. Treating the generated coarse building hypothesis as evidence leads to a set of confidence intervals in image space that can be used as the search space to find the corresponding 2D image primitives and performing a consistency verification of the reconstructed coarse model. Theoretically, in stereo image analysis systems, it is possible to solve the unknown parameters of the 3D model from matches to the homologous 2D image features. However, in practice, the reliability and accuracy of the parameter determination can be substantially improved by fitting the model into the images taken from more than two viewpoints. The methods presented here can be used in either situation. The other important component of the FBMV method is the ability to solve the model parameters by simultaneously fitting all the geometric primitives of the 3D model into the corresponding 2D image features. This is important because it allows the initial matches or the partial matches between the model primitives and the image features to force the location of other structural elements of the model, thereby generating new matches that can be used to verify or reject the initial estimated model parameters.

1.2.2 Achievements

The ultimate goal of this study was to develop a total solution for the automatic extraction and 3D reconstruction of buildings using aerial images in order to fulfill the requirements of a 3D spatial information system. This is achieved through the extension of several existing concepts and developments reported by other researchers in the field of computer vision or digital photogrammetry, as well as introducing novel, mathematically founded concepts and methods, which are developed and implemented (as a prototype) in this study. In this regard, one can classify the contributions of this thesis into two groups, minor and major contributions. The minor contributions are those which are accomplished by extensions of existing methods and are itemized as follows:

1. Development of a least squares planar fit region growing segmentation algorithm. This is the generalization of the segmentation method introduced by Besl (1986).

2. Introducing a two-stage robust parameter estimation method, which is in fact the synthesis and extension of the random sampling type estimators (Fischler & Bolles 1981, Rousseeuw & Leroy 1987), and the M-estimator method proposed by Huber (1981).

3. Development of a new method for computation of adjacency relationships between disjoint objects, in particular polygonal objects with differences in shapes and sizes. This is the extension of the point-wise method originally proposed by Borgefors (1986) in 2D space and extended to 3D space by Chen et al. (1994), and Pilouk et al. (1994).

a b c

Figure 1.2: Three different building roof structures a) gable roof, b) hipped-gable roof, c) complex roof

The major contributions are those which are newly developed in this thesis and have strong roots in theory of parameter estimation and 3D geometry, and are itemized as follows:

1. Introducing a new concept for geometric modeling of generic plane-face solid objects based on the 3D intersection of adjacent polygonal primitives. The concept is implemented as a model generator tool called POLY-MODELER.

2. Introducing a new concept called Feature Based Model Verification (FBMV), for hypothesis validation and modification of polyhedral-like objects. The concept is implemented and applied for the verification of the coarse building hypotheses, generated by POLY-MODELER.

All the proposed methods and algorithms are implemented and the results and their performances are evaluated utilizing real data, and are presented for the whole data set in appendix A. However, for convenience in following up the subsequent processes within this thesis, three different type of buildings, gable, hipped-gable, and complex roof structures (see figure 1.2), are selected from the residential scene of the international Avenches data set (Mason, Baltsavias & Stallmann 1994), and are used through the different chapters of this thesis.

1.3 Thesis Outline

A broad range of topics is covered in this thesis. An overview of the chapters is given in this section.

Chapter 1: an introduction to the problem statement, the motivation of this study and the proposed method and strategy for solving the task.

Chapter 2: a survey covering broadly most previous works on 3D building reconstruction and related techniques described in the literature, and explaining how they fit into the framework proposed in this study. In fact, this is a background chapter and you may wish to skip it, if you are only interested to the development and the contribution of this thesis.

Chapter 3: shortly discusses the principle of the least squares criteria and its shortage to deal with data corrupted by outliers. The concept, properties and general strategy of the major robust techniques for parameter estimation are also reviewed. Finally, a robust parameter estimation method to handle expected outliers in the original observations is developed and the required theory to apply it to the different regression problems encountered during an automated vision process such as 3D object reconstruction is discussed.

Chapter 4: all the low-level processes involved in the recognition part to extract the primary symbolic 2D primitives are described in details. The main emphasis is given to the geometrical characteristics of digital surfaces, extraction of regions of interest, and an iterative region-based segmentation algorithm.

Chapter 5: the subsequent mid-level processes for grouping, structuring and geometric modeling are discussed. This includes the process of transferring intermediate 2D polygonal primitives into the 3D object space, computation of adjacency relationships and merging the compatible adjacent polygons. The second part of the chapter is dedicated to the mathematical concept, notations and a detail discussion of how POLY-MODELER works.

Chapter 6: a brief discussion of the general framework of FBMV method and its internal workflow is given first. The subsequent sections look inside the method and give detail discussions on the fundamental concept of FBMV, its formulation and its robustness. Evaluation of the proposed method, its performance and statistical analysis of the result obtained by some experimental tests conclude this chapter.

Chapter 7: summarizes the contribution and draws our conclusions of this research work and indicates the directions for future research.

Chapter 2

Building Reconstruction in Literature

2.1 Introduction

Recent developments and research in the fields of digital photogrammetry and computer vision for automating the measurement and the scene interpretation tasks have resulted in sophisticated methods, and led to promising results for acquisition and reconstruction of GIS objects. Although a number of research efforts on the extraction of vegetation and forest boundaries are reported (Ebner, Eckstein, Heipke & Mayer 1999), major research efforts are currently put on the detection, extraction, and reconstruction of man-made objects such as roads and buildings captured through passive or active sensors. The major challenge is the automation of the interpretation process, which is usually carried out by operator in the conventional human-based vision system. An automated object reconstruction process requires first an analysis or interpretation of the imagery used, before extraction, structuring, and modeling processes can be performed. Looking at the techniques in the field of digital photogrammetry and computer vision, which are already operational, mainly measuring process of the known or salient points is realized, where no task of interpretation has to be solved. In contrary, the extraction of GIS objects can be regarded as an interpretation problem, because of the variety and complexity of the object of interest. Despite of all the progress that has been achieved in recent years, researchers have not solved the complete automatic acquisition process so far, and there is no fully automatic acquisition system, which could be used, in a wide range of applications. There are some systems that enable the user to acquire buildings automatically, but they can be used only in specific areas of towns, e.g. suburban areas with negligible occlusions and isolated buildings.

A wide range of approaches towards recognition and reconstruction of GIS objects is published in literature (Gruen, Kuebler & Agouris 1995, Gruen, Baltsavias & Henricson 1997, Förstner & Plümer 1997, Förstner, Liedtke & Bückner 1999, Baltsavias, Eckstein, Gülch, Hahn, Stallmann, Tempfli & Welch 1997, Ebner et al. 1999, CVIU 1998). This chapter reviews some of the recent developments reported in the field of semi-automated and automated 3D building reconstruction. The approaches vary a lot in the generality and degree of automation, the used data sources, the geometric modeling techniques and the utilized strategy. Therefore, they can be compared based upon different aspects, and classified based on different criteria such as employed data sources (single/multiple color/bw images, multi-spectral images, laser scanner data, 2D GIS, DSM), supporting object model (specific, generic), amount of user interaction (semi-automated, automated), etc.

Several techniques are discussed from different points of view, taking into consideration different aspects, which are influencing the chain and the result of the reconstruction process. The aim is to demonstrate diversity, and variety of different concepts and approaches, which are proposed and developed for the task of 3D building reconstruction. The discussion is limited to an overview of the proposed acquisition systems, therefore the finer details of the algorithms used in reported applications are not treated here. However, papers describing specific algorithms are cited in the chapters describing parts of the application utilizing similar algorithms, so readers interested in literature describing these algorithms are able to find it there.

Different aspects and criteria, which are influencing the chain and the outcome of the process of recognition and 3D reconstruction of building objects, are discussed in the next section. Some of the methods reported on building reconstruction are picked up from literature and compared in the reminder of this chapter. The main emphasis lies on semi-automated (section 2.3), and automated (section 2.4) approaches.

2.2 Theme in Building Reconstruction

Object reconstruction, in particular buildings, consists of several steps depending on the application, level of required details, primary data sources and cues, methodology used, level of automation and interaction, etc. It mainly consists of detection and features extraction, structuring and grouping, geometric modeling and reconstruction, hypothesis verification, and semantic attribution. All these elements have an impact on the

process of reconstruction and its performance, and yield different types and accuracy of output data. These features can be grouped under several aspects, which are discussed in the following.

2.2.1 Primary Data Sources and Cues

The range of the data, which are used for acquisition and 3D reconstruction of building objects is very large. There is aerial or close range imagery with single, stereo, or multiple image frames. The images are in b/w or color, however multispectral images are also available which are rarely used in reconstruction but often for classification or detection. In addition, there are supporting cues such as DTM/DSM derived directly from laser scanning data or photogrammetric techniques, scanned maps or existing 2D GIS data, beside additional information and knowledge such as position of the sun, time of primary data acquisition, texture, shadows, and reflection properties. The reported algorithms for acquisition of building objects use mainly one or a combination of the above data sources. In fact, depending on the type of basic data, the utilized methods vary. Recently, there is a strong trend towards information fusion and introducing the external and/or a priori knowledge derived from problem domain i.e. 3D geometric constraints, into the acquisition process.

In (Lin et al. 1995) a system is described for detection and describing the buildings from the monocular views of arbitrary aerial scenes. The system uses a perceptual grouping approach (Huertas & Nevatia 1988, Mohan & Nevatia 1989, Huertas, Lin & Nevatia 1993, Collins et al. 1995) to generate the roof hypotheses based upon very specific geometric properties of the building structures, which restrict the shapes of buildings to be a single or composition of rectangular parallelepipeds. The selected hypotheses undergo a validation process based on the shadow (Huertas et al. 1993, Irvin & McKeown 1989, Oddo 1992), and wall verification process (Wang & Schenk 1992). A hypothesis could be validated by either shadow and/or wall evidence, which provide 3D information to the system for creating the 3D model of the building structure. Weidner (1995) reported a different approach still using single DSM cue for acquisition of building objects directly in 3D object space. This approach consists of two main steps. In the first step, building detection is performed based on the grayscale morphological operation (see section 4.4), followed by a reconstruction process where a building model is fitted to the underlying height data. Mass and Vosselman (1999) also reported two methods for extraction of building models from a single high-resolution DSM cue (5 to 10 points per 1 m^2) obtained directly by raw laser scanner data. These methods have the advantage of working on 3D data, which are easier to analyze with respect to buildings, but of course have much less resolution in the ground plane than comparable aerial images. Kulschewski and Koch (1999) reported a method for building recognition based on a dynamic Bayesian network in a single aerial image. The image features, faces of walls and roofs are detected in a face adjacency graph and aggregate to buildings. The Bayesian network is used in order to deal with decisions under uncertainties (Brunn & Weidner 1997, Nevatia et al. 1999).

In order to do object reconstruction, especially buildings, we are interested in 3D information, which can be provided using stereo/multiple frame images. Reconstruction based on geometry of multi frame images is helpful in providing redundant information and improving the reliability of the reconstruction and thus are needed for very accurate measurements (see chapter 6), as well as for identifying occluded areas, or the process of hypothesis verification (Fua & Hanson 1988, Mohan & Nevatia 1989, Ameri 2000). Seen in this light, Nevatia et al. (1999) extended their pervious work utilizing multiple view images both in hypotheses generation and verification process. In (McKeown & McGlone 1993) authors describe a method which uses area and feature based matching techniques in 2D images and fuses the result in 3D object space. Wang and Schenk (1992) introduce a feature based matching technique for interpolation and analysis of urban areas using a 2D edge matching algorithm which generally works on the geometric attributes of the edge such as orientation, length, or extend of overlap (Zong, Li & Schenk 1992, Collins et al. 1995). There are also increasing research efforts in the field of stereo matching performing directly in 3D object space (Haala 1995, Bignone et al. 1996, Baillard et al. 1999), where the matching is performed based on the 2D description of symbolic image features (Brunn, Lang & Förstner 1996), or the multi-view geometry and photometric similarity of the image features e.g., 2D edges (Schmid & Zisserman 1997) over all the images. O'Neill and Denos (1992) describe a couple of problems encountered with stereo matching, such as discontinuities, and shadows.

It is important for the process of reconstruction of buildings to reduce the vast amount of information provided by primary data sources. At the detection level, cues like color and DSM data have proven to be particularly valuable for this task. Most of the proposed methods for building reconstruction focus on region of interest to reduce the search space and to guide the reconstruction process. These cues are used to separate in a first step the man-made structures from the natural ones and then to distinguish building objects from other man-made objects, like roads, bridges, etc. Good success has been reported in the area with isolated buildings (Baltsavias, Mason & Stallmann 1995, Weidner & Förstner 1995, Eckstein & Munkelt 1995, Henricson et al. 1996, Fritsch &

Ameri 1998), however, dense built-up areas still widely resist this approach. Working on DSM derived by airborne laser scanner is reliable and directly supplies 3D coordinates about the surface and is even able to distinguish between the tree canopy and the surface beneath. Laser scanning data is very precise in height but of course has less resolution in the ground plane than comparable aerial images, but yields a higher density of height data. To overcome this problem, there are techniques, which integrate range data from laser scanners and the result from image analysis (Haala 1995). In (Haala 1996) a DSM is used to detect building areas. Instead of applying grayscale morphology he uses height isolines to segment the DSM. The size and the compactness of the segments are computed and those regions, which have building-like attributes are selected for the reconstruction step. Then a stereo matching on straight lines extracted in the image pair is performed, using the height information as approximate values. Taking into consideration that the major ridgeline of the roof structure is mostly elongated in the maximum extension of the building, a building model is fitted to the candidate 3D edges. Lemmens et al. (1997) work on building detection in irregularly distributed laser scanner data sets. He also applies a threshold, which in contrast to previously mentioned methods takes the sensor model of the laser scanner into account. The advantage of this approach is that the sensor characteristics are explicitly modeled and that the original distance measurements are used (Maas 1999).

In the last years there have been a number of research efforts to integrate the background knowledge such as 2D GIS, or existing map data into the process of building acquisition, in particular in detection level. These data can be used very efficiently in those regions in which inherent objects have been acquired already some time before. First the known 2D ground plan extracted from GIS data (Haala & Anders 1997), or derived from scanned map (Nebiker & Carosio 1995) are interpreted and projected into the data set. The projected boundaries of the building in DSM (Nebiker & Carosio 1995, Brenner & Haala 1998*b*, Haala & Brenner 1999), or aerial images (Axelsson 1997) are used as a starting point for 3D reconstruction of building objects.

Due to the complementary properties of different data sources such as DSM, 2D ground plan, and image data, the methods work on data fusion leading to an efficient procedure and appearing comparably straightforward. This is because they naturally require much less model knowledge to solve the problem, and the weaknesses of the individual data set or cue are compensated by the others. Many of the approaches discussed above are already working along these lines, wherever such data, external knowledge, etc. is available.

2.2.2 Supporting Object Model: Specific vs. Generic

The selection of the optimal building model is an essential step for building acquisition, and has to be made before the acquisition step in order to guide the acquisition process. In fact, the goal of the reconstruction process is the determination of the geometric properties and the topological relations between parts of the object, thus requiring as prerequisite some sort of object model (Braun, Kolbe, Lang, Schickler, Steinhage, Kremers, Förstner & Plümmer 1995). To improve the geometric resolution of the utilized model, the number of free structural elements and the number of free parameters have to be increased. Complex models ensure a better fit to real buildings because in real world they vary a lot. However, in small-scale application sometimes a low resolution is favorable. In these cases a generalization process can be performed. Of course, the intended resolution has to be chosen with respect to resolution of the data.

The work on building reconstruction reveals essentially two different basic modeling schemes for the description of the building object, *parametric* and *generic models*. In the case of the parametric object model, the type and relations between model primitives are fixed but their geometry is unknown. This type of model is usually realized in a database of predefined building types, or simple volumetric building primitives which are matched with the corresponding features in images or cues such as a DSM to estimate the correct geometry of the model primitives and reconstruct the building instance. In the case of the generic object model, the numbers of model primitives as well as their geometry and topological relations are unknowns. This type of model allows for variation in the structure of the object model, thus they need a mechanism of specifying the internal structure of the object based on a set of geometric parameters that is not fixed in number. The buildings are reconstructed by geometric grouping of extracted image features. Thus its quality and the degree of completeness depends heavily on the result of the feature extraction process, which deals in most cases with the extraction of the contour lines and/or the homogeneous surfaces of the buildings in the scene. Whereas the former approaches yield both geometric and to a certain extent semantic information of 3D buildings, their fixed number of predefined buildings, or building parts limits their application. Indeed the practical use of parameterized models depends on the capacity of the assembled building types. Urban scenes show irregular man-made structures and very complex combination of buildings and building parts. These buildings and building formations cannot be modeled with this type of models. In contrast, the generic model of the latter approaches allows the representation of the arbitrarily shaped buildings, but provide no object specific interpretation of the reconstructed building. The major problem

of generic models is that buildings can only be reconstructed completely, if the feature extraction process detects all the significant primitives of the building structure. As it does not take into account any a priori knowledge about the building, thus, there is no prediction about the missing or occluded parts of the building. To circumvent such conditions multiple images taken from different viewpoints for hypothesis generation, or verification process can be used (Faugeras et al. 1995, Henricson et al. 1996, Fischer et al. 1998, Hendrickx et al. 1997, Baillard et al. 1999, Ameri 2000).

As it discussed above, the type of modeling used to express the building objects in the scene limits the class of buildings to be recognized. In (Fua & Hanson 1988, Huertas & Nevatia 1988, Irvin & McKeown 1989, Mohan & Nevatia 1989, Price & Huertas 1992, Shufelt & McKeown 1993, Nevatia et al. 1997) building objects are modeled as rectangular shaped blocks with vertical walls. This approach is extended to flat-roof arbitrary shaped prismatic model (Shi, Shibasaki & Murai 1997, Weidner 1996), and non-flat roof parametric model (Mueller & Olson 1993, Haala 1995, Spreeuwers, Schutte & Houkes 1997, Maas 1999). A disadvantage with these ways of modeling is that it is very difficult to recover the correct structure of the buildings, which have complex structures. This is due to the limited number of predefined building types, thus only buildings which are explicitly stored in the database can be reconstructed.

To model complex buildings a generic object model is required. This type of model can be realized either in the form of constructive solid geometry (CSG) model, or in the form of a boundary representation (b-rep). The CSG model is an aggregation of a set of parameterized volumetric primitives along with given relation rules which could be a Boolean operation such as intersection, union, or subtraction (Englert 1997). The parameterized volumetric primitives are building models with fixed topology and variable size. Modeling buildings by volumetric primitives has several particular advantages. First, as every building type is explicitly modeled, their different forms of appearance can be derived a priori. Second, even partially occluded buildings can be fully reconstructed. However, the identification of an instance of a volumetric primitive in the scene required a priori knowledge, or a hint by the operator. The other shortage is the lack of flexibility with respect to different building shapes, in particular when there are irregularities in building structure. Parameterized volumetric primitives are employed by, among others (Lang & Förstner 1996, Jaynes et al. 1997, Brenner & Haala 1998*b*, Haala & Brenner 1999, Gülch, Müller & Läbe 1999).

Polyhedral models in the form of a b-rep have shown to be adequate and most flexible supportive generic object model for building reconstruction (Henricson et al. 1996, Hendrickx et al. 1997, Ameri & Fritsch 1999). There is no restriction on the form of the building, except the planarity of the surfaces. Obviously this is a quite general model, just excluding curved surfaces, thus allowing various roof structures for representing arbitrarily shaped buildings.

As it mentioned above, the reconstruction methods based on the generic building models are heavily dependent on the low-level geometric primitives extracted from the scenes. As these are so far very hard to derive from the complex scenes such as aerial images, detection and extraction of image feature is regarded as a bottleneck for object reconstruction.

2.2.3 Methodology

The discussion on methodologies applied in building acquisition is categorized into the three topics of *work flow*, *feature extraction*, and *structuring and grouping*. This is due to the fact that each unit has a significant impact on the final methodology used to solve the problem.

Work flow: From a work flow point of view, there exist two basic approaches, bottom-up, and top-down processes. Bottom-up process is a data-driven strategy, which extracts in a first step image primitives such as points, edges, and/or homogeneous regions, groups them to higher level entities and through the process of hypothesis generation, builds up the complete object. The main problem here is the instability of the segmentation process at the lowest level, which is mostly caused by the presence of the noise in the input data, and the ambiguity of higher level process of grouping, which is normally controlled using domain specific knowledge. Man-made objects such as buildings represent structures that are not random but have specific geometric properties. Those properties can be used to organize the extracted features or image primitives into roof and building hypotheses. There exists a multitude of techniques that take advantage of different kinds of knowledge about the object in order to generate building hypotheses such as probabilistic relaxation (Heuser & Liedtke 1990), Bayesian reasoning (Kulschewski & Koch 1999, Brunn & Weidner 1997, Nevatia et al. 1999), constraint satisfaction networks (Mohan & Nevatia 1989, Price & Huertas 1992), geometric (Wiman & Axelsson 1996, Hendrickx et al. 1997), and semantic reasoning (Lang 1999).

The top-down approach is model-driven and usually starts with extracting features, followed by matching them to a library of stored objects (Jaynes et al. 1997, Stilla et al. 1997, Spreeuwers et al. 1997). Various geometric constraints help to reduce the search space thus keeping the combinatorial explosion of searching under control. Essential to this technology is the object model itself, which is often used in explicit form. The object data structure inferred from the image(s) is matched to the model structure. While this concept has a certain justification in robotics and navigation, where the environment might be of reduced complexity, we encounter some problems where it comes to recognizing objects with complex shapes in outdoor scenes. Even rather simple structures, such as buildings, come in such a variety of different sizes and shapes that it is a fruitless attempt to precisely describe and store all of them in a model library (Schenk 1993).

In more recent approaches of building extraction we see elements of both strategies used together in an interrelated manner (Henricson et al. 1996, Gruen et al. 1997, Fischer et al. 1999, Ameri & Fritsch 1999). This seems to be the right way to approach the problem. Following this approach, first hypotheses are generated about the existence of objects in the scene. These are later verified with a robust verification mechanism (Fua & Hanson 1988, Lowe 1991, Zhang, Sullivan & Baker 1992, Ameri 2000).

Feature extraction: The process of object acquisition or reconstruction is usually initialized by the detection of the object of interest e.g., building, in the scene. This is performed by a coarse segmentation or classification of the scene to the regions, which have the potential of being an object of interest. In fact this is an application-dependent scene interpretation process (Schenk 1993), where the problem-domain knowledge is used to reduce the dimensionality of the search space, and thus the amount of the raw data which has to be analyzed. This is in contrary to the general domain-independent scene interpretation system (Brooks 1981), where the reasoning engine is independent of the scene type studied. The detection process and extraction of geometric primitives are technically different in 2D intensity images, and in 3D cues such as a DSM. Detection of buildings can rely on simple attributes that distinguish buildings from non-buildings. In DSM this might be the relative height and the size of regions with heights larger than the surrounding (Haala 1995, Baltsavias et al. 1995, Weidner & Förstner 1995, Eckstein & Munkelt 1995, Henricson et al. 1996, Fritsch & Ameri 1998). In images the situation is much more complicated due to the loss of the third dimension. In fact the extraction of relevant low level image primitives from a complex scene such as an aerial image is a complex procedure and its complexity is increased with a decrease in the dimension of the image primitives (see section 4.2). This operation is usually regarded as a bottleneck process in automated building acquisition systems due to several reasons. Low contrast between the roof and the surrounding area causes the low-level segmentation to be fragmented. In addition cars and neighboring trees cause further fragments and noisy borders.

Several methods have been reported which extract different image primitives to initiate the reconstruction of building objects in the scene. There are approaches which extract corner points with long neighboring image edges for 3D reconstruction of corners of buildings (Lang 1999), the methods which use image edges directly to setup building hypotheses (Collins et al. 1995, Nevatia et al. 1999), or those methods which first extract 3D edges from the images to group them to the planar patches of the building roof structure (Henricson et al. 1996, Hendrickx et al. 1997, Baillard et al. 1999), or the methods which directly extract the planar polygons (Ameri & Fritsch 1999, Fradkin, Roux & Maitre 1999), and the one which use the combination of points and edges primitives (Jaynes et al. 1997), as well as regions (Fischer et al. 1999).

Structuring and grouping: Structuring and grouping of the data is another difficult part of a 3D building reconstruction process, in particular when a generic model is used. Structuring essentially is setting up the neighborhood relations, e.g., the topology between the different parts of a building, organizing and representing them in a suitable form in order to generate building hypotheses or an image model being the projection of the building model.

Instantiating the building models by extracting distinct features from single image and grouping them on the sole basis of 2D geometry (Huertas & Nevatia 1988, Price & Huertas 1992) is bound to be combinatorially explosive since the 2D geometry alone does not sufficiently constrain the 3D reconstruction problem. In addition, derivation of 3D structures of the object from one image is not unique, as an image is a 2D projection of the 3D real world object and it therefore contains only a part of the object information. Therefore, it is necessary to incorporate information from other sources of data e.g., DSM, or using multiple or at least two images (Fua & Hanson 1988, Mohan & Nevatia 1989, Nevatia et al. 1999). In recent years, the use of 3D information has emerged as a powerful means to disambiguate complex scenes, since the expressiveness of 3D data is higher that that of 2D data. A strategy being traced by many researchers is an early transition from 2D to 3D in the reconstruction process, this way reducing the overall number of future hypotheses. This is done by extracting meaningful features in the image, which have correspondences to building primitives, such as 2D points corresponding to 3D building corners (Fischer et al. 1998, Faugeras et al. 1995), 2D edges correspond to 3D edge structure of the building (Haala 1995, Baillard et al. 1999, Bignone et al. 1996), or 2D planar regions correspond to 3D planar-

patches of roof structure (Fradkin et al. 1999, Ameri & Fritsch 1999). The extracted 2D features are transferred to 3D objects using 3D stereo matching techniques, or are based on the approximate terrain surface model, thus the geometric modeling and structuring consistently is performed in 3D, imposing the 3D geometry constraints derived from problem-specific knowledge (Henricson et al. 1996, Jaynes et al. 1997, Haala et al. 1998, Ameri & Fritsch 1999, Baillard et al. 1999, Fischer et al. 1999). Note that mutual interaction of 2D and 3D processes is required at all levels of reconstruction in particular the structuring phase. Whenever 3D features are incomplete or entirely missing, additional 2D information should be used to infer the missing features and structures.

2.3 Semi-Automated Methods

It is discussed above that the 3D building reconstruction consists of several processes mainly detection, extraction of distinct feature, structuring, and geometric modeling. The concept of the semi-automatic approach is that the operator performs the interpretation, while the measuring tasks are performed by the system as far as possible. Interpretation can be regarded as detection and/or structuring, which usually takes place in one single image/cue. The automatic measurement then may use multiple images for feature extraction and geometric reconstruction depending on the utilized methodology. Actually, the goal is to increase performance of purely operator driven systems by inclusion of automatic techniques and assisting the operator from doing structured measurements. The semi-automatic method aims at taking the advantage of both, the operator's skills to interpret the data and controls the acquisition process, and the machine's skill to efficiently handle large amount of data and accelerate the measurement process, thus, achieving higher productivity in acquisition.

Semi-automatic systems can make intensive use of automated procedures. For example, automatic extraction of geometric features, e.g., 2D corner points or edge segments, assist the operator to do the measurement right in the extracted primitives, thus reducing the operator's task to pointing at the correct image features or to provide accurate enough approximate values. Automatic stereo matching techniques enable automatic measurement of heights for the extracted 2D features by finding corresponding features in two or more images. Automated matching of model primitives to their homologous features in the image(s) allows automatic determination of model parameters, to mention a few.

Several systems have been developed in the last years. Three systems, among others, are selected and discussed in this section, which are operational and productive. The selected systems are presented and reported in several publications, while this section only refers to the recent publication.

CC-Modeler (Gruen 1999): CC-Modeler is a semi-automated 3D object reconstruction system. The feature identification and measurement of the 3D point clouds is implemented in a manual mode, on an analytical plotter or a digital station. During the data acquisition, 3D points belonging to a single object should be coded into two different types according to their functionality and structure, boundary or interior points. Boundary points must be measured in a particular order, either clockwise or counter-clockwise. Interior points can be measured in an arbitrary sequence. Since the human operator is responsible for the interpretation and measurement, it is possible to acquire any level of object details for buildings, roads, waterways, and other objects, which may be approximated by a polyhedral model. The next step is fitting planar structures to the measured sets of point clouds. A Consistent Labeling algorithm implements this by probability relaxation operations. As the results of Consistent Labeling, CC-Modeler delivers the face definition for every face. Then, a least squares adjustment is performed for all faces simultaneously, fitting the individual faces in an optimal way to the measured points and considering the fact that individual points may be member of more than one face. This adjustment is amended by observation equations, which model orthogonality constraints of pairs of straight lines. Finally, a vector description of 3D objects is obtained. Although the procedure of geometric modeling is automated, human intervention and interaction with the automatic procedures is also available. The CC-Modeler has been tested successfully in several projects.

ObEx (Gülch et al. 1999): The reported system works on the principle of constructive solid geometry (CSG) for the modeling of complex buildings, the operator guidance, the assistance of automated modules to perform a certain number of measurement tasks, and utilizing multiple images. Within this approach buildings are reconstructed by combining a series of 3D volumetric primitives until the complete building has been modeled. In addition, for reasons of efficiency three different parameterized building models are stored in the system as well. The task of the operator is to choose a primitive which will be projected as a wire-frame model into the focussed building in a single image, and adjust the parameter of the model. The adoption of the model parameters can be done in three different ways, 1) a purely manual adoption based on a series of point measurements, 2) a guided adoption using a priori extracted image edges, and 3) an automated adoption based on the automatic correlation and matching techniques. The operator is, however, at all stages in the position to interact and

perform a purely manual fitting of the models. The method has been tested in several projects. It works very well in suburban areas and moderate building roof structure, however, for modeling very complex buildings in downtown areas, with many small details, some unsolved problems are reported which are mostly inherited due to the limitation of CSG modeling to describe the geometrical structures in details. In fact, they fail to represent the irregularities in the building structure, which are nowadays increasing due to the new styles being developed or invented.

Ifp method (Haala & Brenner 1999): This system uses a high resolution DSM, mostly derived directly by laser scanning, and 2D ground plans as basic data sources. Since there is not necessarily any interaction during the reconstruction process, this method can be categorized as an automated system as well. However, as the process uses the digital 2D ground plans, which is acquired or digitized manually, there are limits to be overcome by semi-automated editing. The method also follows the CSG modeling paradigm, a complex building is reconstructed by combining its basic units (parameterized volumetric primitives). The process starts by decomposing the polygonal ground plan of the building into the 2D rectangle primitives. Each 2D primitive is the footprint of a corresponding 3D primitive. The location, orientation and the size of the 2D primitive applies for the 3D primitive as well. The remaining parameters of the roof structure such as roof type and its slope, as well as height of the building are determined in a least squares estimation process, which fit the model to the underlying DSM. When several models are suitable, the one with the smallest residual is selected. In order to provide a visual control on the reconstructed building and to allow the manual refinement and modification of the building models an efficient editing tool is also implemented. The system has been successfully tested on a number of large projects, and achievement of very high reconstruction rate is reported.

2.4 Automated Methods

Automatic techniques for building extraction have evolved rapidly in the last years. They show great potential for 3D reconstruction of building objects. Moreover, due to the higher amount of data required for 3D data acquisition and the need for generating topologically consistent description of the object, they appear to be the only way to satisfy the needs of the users (Förstner 1999). However, up to now they are not reliable enough to be used in practice. They are mostly in experimental level and new development is on the way. The reported techniques vary depending on e.g., feature segmentation and grouping, or geometric modeling either in 2D or 3D, and have been mainly applied to suburban areas with isolated buildings without or little vegetation in their vicinity.

Two systems, among others, are selected and discussed in this section. The selected systems are also presented and reported in several publications, but in this section only the recent publication is referred to.

ARUBA (Henricson 1998): ARUBA utilizes multiple color aerial images, and a generic polyhedral models to support the reconstruction process. The process starts with extraction of regions of interest by detecting the elevation blobs from the DSM and combining this information with color analysis. The fact that each region of interest may include only one building simplifies the automatic reconstruction to a large extent. The general assumption is that a complete roof consists of a set of planar parts, which mutually adjoin along their boundary. In a first step, straight 2D edges from one image are matched to the corresponding edges in other images, thus producing 3D edges, which are then grouped to plane hypotheses. The object boundary of each plane hypothesis is then found by extracting 2D enclosures employing a new grouping technique, which is based on similarity in proximity, orientation, and photometric and chromatic region attributes. The most evident and consistent set of planar roof hypotheses is finally selected based on simple geometric criteria. Vertical walls are added afterwards by projecting the eaves of the roof down to the underling DTM. The final result is a complete CAD model of the roof and its vertical walls, including their topological relations.

Bonn method (Fischer et al. 1999): This method uses multiple aerial images and a CSG modeling scheme, where the primitives are building parts, representing either end parts of basic buildings or connecting parts between pairs of basic buildings. The process starts with extraction of 2D image features such as points, edges, and regions. The extraction process leads to the 3D reconstruction of corner points, which are promising features for generating building part hypotheses. Building parts are classified by their roof type. A strongly model-driven aggregation combines 3D local building parts to more complex 3D building aggregates. The resulting 3D building hypotheses and their components are projected into the images to allow a component-based and robust hypothesis verification applying constraint-solving techniques.

Chapter 3

Robust Parameter Estimation

3.1 Introduction

Given a mass of qualitative observations, e.g. grey value pixels which are normally ordered in a digital image format, an automated image interpretation system intends to summarize and describe the data by a series of quantitative geometric primitives or an instantiation of an appropriate model. The basic approach of estimating the model parameters is usually the same. A *cost* or *merit* function that measures the agreement between the observations and the hypothesis model is defined. The cost function is conventionally arranged so that small values represent close agreement. Accordingly, the parameters of the model are adjusted and estimated in order to achieve a minimum cost, thus yielding best-fit parameters, in other words solving an optimization problem. There are also other important issues in this process that go beyond the problem of finding best-fit parameters. In fact, observations are generally not exact and they are subject to measurement errors which in the context of computer vision are called *noise*. Thus, they never exactly fit to the model, even when the model is correct. Therefore, tools to assess whether or not the model is appropriate, to measure the quality of the fit, and to estimate the accuracy of the derived parameters are required. In addition, it is often the case that the cost function has many local minima, whereas the best result of the optimization is obtained when the global minimum is found among the many local minima.

Traditionally, standard least squares (LS) framework has been used for regression analysis or model fitting. In fact, it is the basis for many estimation procedures which attempt to minimize the cost of the errors in the estimate (Mikhail 1976). Although LS criteria may be optimal and reliable if underlying noise in the original measurement can be modeled as a Gaussian distribution, it is not optimal for other type of noise distribution. If the noise distribution includes outliers, the distribution is heavy-tailed and LS criteria can lead to very poor estimators (Fischler & Bolles 1981, Rousseeuw & Leroy 1987, Schunck 1990). In such circumstances, utilization of a *robust parameter estimation method* is esential. There are classes of computations in the field of *robust statistics* that have been designed to handle outliers (Huber 1981, Hampel et al. 1986, Rousseeuw & Leroy 1987, Förstner 1987). These methods are currently gaining popularity in the field of computer vision and have been applied in the number of vision procedure (Bolles & Fischler 1981, Haralick & Joo 1988, Lee, Haralick & Zhuang 1989, Sester & Förstner 1989, Kumar & Hanson 1990, Sinha & Schunck 1989, Schickler 1992, Axelsson 1996, Torr & Zisserman 1997, Gülch et al. 1998). The main contribution of this chapter is to design a – or to extend the existing – robust parameter estimation to handle expected outliers in the original observations and develop the required theory to apply it to the different regression problems encountered during an automated vision process such as 3D object reconstruction. In particular in this study the proposed method is used to transfer extracted 2D plane-roof polygons in image space into the 3D plane-roof polygons in object space by fitting the 2D primitives over the existing DSM. This process is elaborated in more detail in section (3.4).

The next section shortly discusses the principle of the least squares criteria and its shortage to deal with data corrupted by outliers. The concept, properties and general strategy of the major robust techniques for parameter estimation is discussed in the subsequent section. Finally, a synthesis, two-stage robust parameter estimator developed based on a *random sampling* type estimator (Fischler & Bolles 1981, Rousseeuw & Leroy 1987) complemented by the robust estimation technique *M-estimator* proposed by (Huber 1981) is introduced. The ability to couple these robust techniques enables us to arrive at an empirically optimal, and statistically satisfying method. The technique used and conclusions drawn have applicability to the broad range of computer vision problems troubled by outliers.

3.2 Least Squares Principles

Computer vision algorithms generate interpretations of the observed data. These algorithms are typically cast in terms of the minimization of an appropriate cost function, and in many cases this cost function is the sum of

squares of a set of residuals, the *least squares solution*. This is usually for the reason that LS is the *maximum likelihood estimator* when the observation error is Gaussian. Consider fitting a set of n observed points (x_i, y_i), $i = 1, ..., n$ to a model that has u adjustable parameters a_j, $j = 1, ..., u$. The model predicts a functional relationship between the measured independent and dependent parameter variables as follows:

$$y_i(x) = y(x_i; a_1, a_2, ..., a_u) \tag{3.1}$$

where the dependence on the parameters is indicated explicitly on the right-hand side. According to LS criteria, the best parameters are obtained by minimizing the following cost function:

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} ((y_i - y(x_i; a_i, ..., a_u))^2 \ \rightarrow \ minimize. \tag{3.2}$$

Now, suppose that each data point y_i has a measurement error that is independently random and distributed as a Gaussian distribution around the true model $y(x)$. In addition, the standard deviation σ of the error distributions are the same for all the points. Then the probability p of the data set is the product of the probabilities of each point and it can be expressed as:

$$p = \prod_{i=1}^{n} \{\exp\left(-\frac{1}{2}(\frac{y_i - y(x_i)}{\sigma})^2\right) \Delta y\}. \tag{3.3}$$

Maximizing equation (3.3) is equivalent to maximizing its logarithm, or minimizing the negative of its logarithm, that is :

$$-\log p = \left(\sum_{i=1}^{n} \frac{(y_i - y(x_i))^2}{2\sigma^2}\right) - n \log \Delta y \tag{3.4}$$

since n, σ, and Δy are all constants, minimizing this equation is equivalent to minimizing equation (3.2). The above formulation expresses that the least squares fitting is a maximum likelihood estimation of the fitted parameters if the measurement errors are independent and normally distributed with constant standard deviation. However, when outliers contaminate the data, this justification no longer holds. Outliers, which are inevitably included in the original measurements can distort a fitting process in such a way that the resulting fit can be arbitrary. This is illustrated in figure (3.1), taken from Fischler and Bolles (1981), which shows the result of the linear regression using two different estimators on a data set, which contains a gross error, *point 7*.

The least squares estimator provides an erroneous solution, *fit 2*, whereas their proposed *RANSAC* robust estimator, which will be described in next section, gives a solution, *fit 1*, that well fits the six inliers. This data set demonstrates the failings of the standard least squares and heuristic attempts to remove outliers. Disregarding the point with largest residuals after LS fitting, remove *point 6* not *point 7*. Indeed repeated application of this heuristic method to convergence results in half the valid data being discarded, and *point 7* remaining as an inlier to a completely erroneous fit.

The physical analogy shown in figure (3.2) may make this discussion more clear (Schunck 1990). Given a set of points in the plane, the center of the mass is required. *"Attach springs with equal constants to the fixed points and to a small object that can move freely. The object will be pulled to the average of the locations of the points. The springs implement a least squares norm through the spring equation for potential energy. This physical analogy corresponds to the derivation of the calculation of an average from the criterion that the sum of the squares of the residuals, the difference between each point and the average, should be minimized. Now suppose that one of the points can be moved. Call this point a leverage point. It is possible to force the location of the average to be shifted to any arbitrary point by pulling the leverage point far enough away. This illustrates the extreme sensitivity of the estimators based on least squares criteria to outliers. Even a single outlier can ruin an estimate"*(Schunck 1990, pp. 4). Because of the lack of robustness efforts have gone into producing estimation methods that are more robust than least squares.

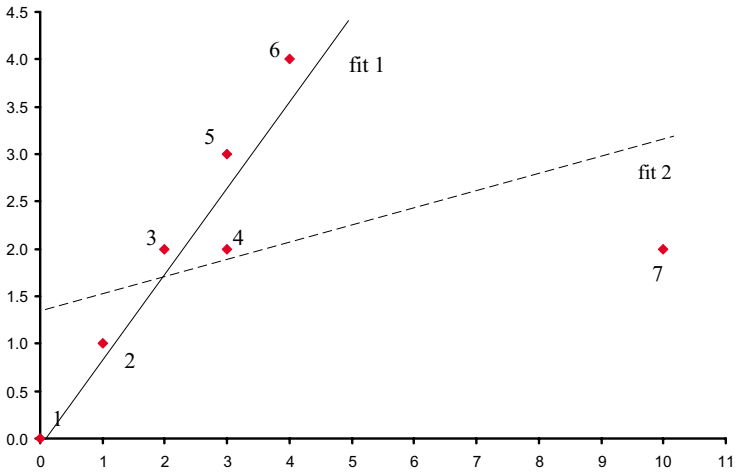


Figure 3.1: Linear regression: fit 1) six of the seven points are selected as inliers and the best fitted line are obtained by RANSAC (courtesy of Fischler and Bolles 1981), fit 2) least squares estimation provides an erroneous solution

For example, changing the spring constants in such a way that the points which are far away have little influence on the estimated parameters, this is equivalent to the idea of giving less weight during the estimation process to a point with a larger error (Huber 1981, Hampel et al. 1986). Or alternatively, breaking the springs attached to the outliers so that the estimate remains unharmed which is the concept of a random sampling method (Fischler & Bolles 1981, Rousseeuw & Leroy 1987).

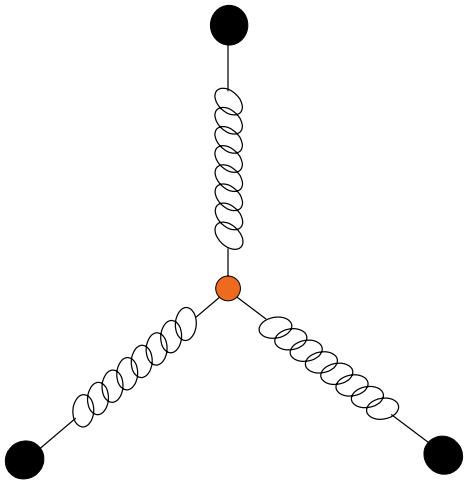


Figure 3.2: Physical analogy that illustrates the sensivity of least squares methods to outliers (courtesy of Schunck 1990)

The principle of LS estimation is the essential part of many robust techniques. In addition, its theoretical concept complemented by statistical test theory provides a comprehensive framework for assessment and evaluation of the estimated results as well as improvement of the measuring configuration when the underlying noise model is normally distributed. Therefore, because of its importance and in order to complete our discussion, a short overview of the quality analysis of the least squares estimation and its quality measures is given. A detail discussion of this topic can be found in (Baarda 1967, Baarda 1968, Förstner 1987, Koch 1999).

Let the linear model (3.5) with the assumption (3.6) be given

$$E(\mathbf{l}) = \mathbf{A}\mathbf{x} \quad \Rightarrow \quad \mathbf{l} + \mathbf{v} = \mathbf{A}\mathbf{x} \tag{3.5}$$

$$D(\mathbf{l}) = \sigma_0^2 \mathbf{P}^{-1} \tag{3.6}$$

where

- $\mathbf{A}$ is the associated design matrix including partial derivatives of the observations with respect to the unknowns,
- $\mathbf{x}$ is the vector of u unknown,
- $\mathbf{l}$ is the vector of n observations
- $E(\mathbf{l})$ is the expection of the observations
- $\mathbf{P}$ is the corresponding weight matrix,
- $D(\mathbf{l})$ is the dispersion or the variance-covariance matrix of the observations,
- $\mathbf{v}$ is an added random errors vector, and
- σ_0^2 is an unknown variance factor.

The system of (3.5) is the well-known Gauss-Markov model. The least squares principle in this model leads to estimates $\hat{\mathbf{x}}$ for the unknowns, $\hat{\mathbf{v}}$ for the residuals, and $\hat{\sigma}_0^2$ for the variance factor according to following formulation:

$$\hat{\mathbf{x}} = (\mathbf{A^T P A})^{-1}\mathbf{A^T P l} \tag{3.7}$$

$$\hat{\mathbf{v}} = \mathbf{A}\hat{\mathbf{x}} - \mathbf{l} \tag{3.8}$$

$$\hat{\sigma}_0^2 = \frac{\hat{\mathbf{v}}^\mathbf{T}\mathbf{P}\hat{\mathbf{v}}}{r} \tag{3.9}$$

where $r = n - u$ is the *redundancy* of the system. Quality aspects of LS estimation basically refer to *precision* and *reliability*. The analysis of the precision of the estimates is mainly based on the variance-covariance matrix of the result which reflects the influence of random errors in the observations onto the estimated parameters and can be derived by the error propagation law as follow:

$$\hat{D}(\hat{\mathbf{x}}) = \hat{\sigma}_0^2(\mathbf{A^T P A})^{-1} \tag{3.10}$$

$$\hat{D}(\hat{\mathbf{l}}) = \mathbf{A^T}\hat{D}(\hat{\mathbf{x}})\mathbf{A} \tag{3.11}$$

$$\hat{D}(\hat{\mathbf{v}}) = \hat{D}(\hat{\mathbf{l}}) - D(\mathbf{l}). \tag{3.12}$$

The reliability analysis [1] uses reliability measures such as *sensitivity or robustness factor, contribution numbers*, and *redundancy numbers*, which determine the maximum influence of undetectable errors in the observations onto the estimates, the contribution of each observation onto the determination of the unknown parameters, and how the model errors show up in the corresponding residuals, respectively. The later measure, redundancy numbers, is used in our synthesis robust parameter estimation method (section 3.4) for the detection of outliers and computing the standardized residuals. The quality analysis of the LS estimation as discussed above complemented by the available statistical test under the assumption of a Gaussian noise model can be used as a mechanism for quality control of the vision tasks as well as the planing purposes and design of the measurement configuration.

[1] According to Baarda (1968), one can distinguish between *i) internal reliability*, a quality measure with respect to the delectability of the model errors, which define a lower bound for detectable errors, and *ii) external reliability*, a quality measure with respect to the sensitivity or robustness of the result, which uses relations describing the maximum influence of undetectable errors onto the estimates.

3.3 Robust Parameter Estimation Methods in Computer Vision

An automated vision process such as object reconstruction requires not only describing the geometry of the object of interest but also the ability to deal with incorrect data which will inevitably arise in a real system. They must be able to interpret the data while simultaneously reject the gross errors, called outliers. These are data points with large errors that do not agree with the postulated error model, in other words, they do not fit the assumed error distribution. Many image analysis procedures assume that errors in the original observations have a Gaussian distribution or are normally distributed. The Gaussian distribution does not have broad tails, which means that most errors are small and are concentrated around the center of the distribution. As discussed in the previous section, standard LS methods which are popular in many areas of science are not robust to violations in problem assumptions. There are classes of computations in the field of robust statistics that have been designed to be robust to wide deviation from the assumptions. They are able to perform when the assumptions underlying the estimation, say the noise model, are not wholly satisfied (Huber 1981, Hampel et al. 1986).

There are two important measures used by robust statistics and vision community to evaluate the robust algorithms. These are the *statistical* or *relative efficiency* and *breakdown point*. The breakdown point is the smallest fraction of outliers present in the original data that may cause the output estimate to be arbitrarily wrong. In other words, it is the largest percentage of the outliers that can be tolerated by the estimation algorithm before the breakdown occurs. Therefore the higher this percentage, the better. For a standard LS estimate one outlier is sufficient to alter arbitrarily the result, therefore it has a breakdown point of $1/n$, where n is the number of points in the set. An indication of the breakdown point is gained by conducting the tests with varying proportions of outliers (Rousseeuw & Leroy 1987). Statistical efficiency is the ratio between the lowest achievable variance for the estimated parameters and the actual variance provided by the given method, so that the best possible value is 1. It is the ability of the algorithm to correctly recover the characteristics of the original data and is the traditional measure used to evaluate a fitting process. In the presence of a Gaussian noise distribution, the LS estimation is known to be the most statistically efficient estimator (Kim, Kim, Meer, Mintz & Rosenfeld 1989). In fact, there is always a trade-off between algorithms with high breakdown points versus those with high efficiency.

The reminder of this section is a general overview of the most common robust estimation methods used in the fields of statistics and vision. It is not our intention to perform a detail comparative study based on the efficiency and robustness of these methods, as it is beyond the concern of this research work. Thus, only a general discussion of the concept and basic strategy of every method is given with some references to their applications reported in the field of vision. However, two well-known techniques of *M-Estimator* and *Random Sampling* are elaborated in more detail, as they are the fundamental parts of our proposed two-stage synthesis robust estimation technique (section 3.4).

3.3.1 M-Estimator

The M-estimation techniques have been developed by Huber (1981), and follow the maximum-likelihood formulations in order to derive the optimal weighting for the data under non-Gaussian conditions. In contrast to least squares criteria, which minimize the sum of squares of an error function (equation 3.2), the M-estimator techniques minimize the sum of a function $\rho(d_i/s)$, where d_i is the error function for the data point i and s is a scaling factor. In other words, the parameters that minimize equation (3.13) are sought.

$$\sum_{i=1}^{n} \rho(d_i/s) \;\rightarrow\; minimize \tag{3.13}$$

The form of ρ is derived from the particular chosen density function in the manner similar to the case of Gaussian error function and it should satisfy the following assumptions:

1. It is a continuous function and has a unique minimum, $\rho(0) = 0$,
2. It is a symmetric function, $\rho(u) = \rho(-u)$,
3. It is a positive function, if $0 \leq u \leq v$ then $\rho(u) \leq \rho(v)$,
4. It is a definite function, if $a = \sup\ \rho(u)$ then $0 < a < \infty$, and
5. It is an increasing function, if $\rho(u) < a$ and $0 \leq u < v$ then $\rho(u) < \rho(v)$.

There are many different minimum functions ρ proposed in the literature (Huber 1981, Hampel et al. 1986), which have been applied in the number of vision problems (Förstner 1986, Haralick & Joo 1988, Lee et al. 1989, Sester & Förstner 1989, Kumar & Hanson 1990, Sinha & Schunck 1989, Schickler 1992, Wild, Krzystek & Madani 1996, Torr & Zisserman 1997). Usually the density function is chosen so that ρ is some weighting $\rho(d_i/s) = w_i d_i$, of the error that reduces the effect of the outliers on the estimated parameters. In fact, this is motivated from the concept of *influence function* or *influence curve* proposed by (Hampel 1968, Hampel et al. 1986). The influence function is an heuristic interpretation tool that describes the effect of an outlier at a data point i on the estimate. It is defined as a function proportional to the first derivative of the minimum function:

$$\psi(x) = \partial\rho(x)/\partial x. \tag{3.14}$$

Let a_j represent the set of unknown parameters to be estimated. Differentiating the cost function expressed in equation (3.13) with respect to each parameter a_j, we get the set of equations:

$$\sum_{i=1}^{n} \psi(d_i/s)\frac{\partial d_i}{\partial a_j} = 0 \tag{3.15}$$

where the influence function ψ is the derivative of the ρ with respect to errors d_i. The equation (3.15) can be written in the standard weighted form as:

$$\sum_{i=1}^{n} w_i d_i \frac{\partial d_i}{\partial a_j} = 0 \tag{3.16}$$

where $w(x_i)$ is called *weight function* and is defined as:

$$w(d_i/s) = \frac{\psi(d_i/s)}{d_i/s}. \tag{3.17}$$

The influence function of a Gaussian noise model with the minimum function $\rho(x) = \frac{1}{2}x^2$ also confirms the non-robustness of the least squares estimation method. The $\psi(x) = x$ indicates that the more deviant the points, the greater its influence.

A typical minimum function scheme $\rho(x)$, proposed by (Hampel 1968) corresponds to

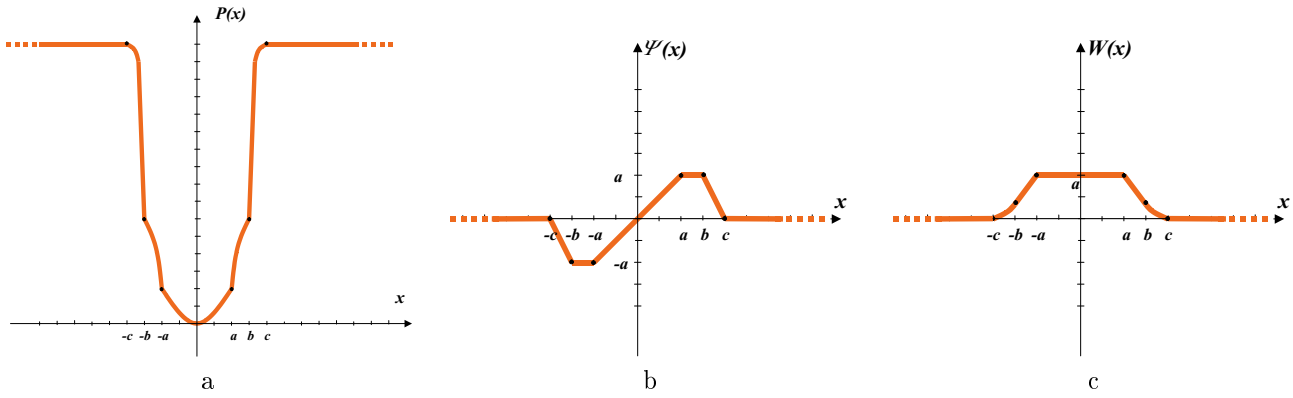
$$\rho(x)_{a,b,c} = \begin{cases} \frac{x^2}{2} & 0 \le |x| \le a \\ a|x| & a \le |x| \le b \\ \frac{a(c|x| - \frac{x^2}{2})}{c-b} & b \le |x| \le c \\ \frac{ac^2}{2(c-b)} & c \le |x| \end{cases} \tag{3.18}$$

and is called *Hampel three-part* minimum function, where $0 < a < b < c < \infty$. The corresponding influence $\psi(x)$ and weight $w(x)$ functions are formulated in the equations (3.19) and (3.20) respectively.

$$\psi(x)_{a,b,c} = \begin{cases} x & 0 \le |x| \le a \\ a\ sign(x) & a \le |x| \le b \\ \frac{a(c-|x|)}{c-b} sign(x) & b \le |x| \le c \\ 0 & c \le |x| \end{cases} \tag{3.19}$$

$$w(x)_{a,b,c} = \begin{cases} 1 & 0 \le |x| \le a \\ \frac{a}{|x|} & a \le |x| \le b \\ \frac{a(c-|x|)}{(c-b)|x|} & b \le |x| \le c \\ 0 & c \le |x| \end{cases} \tag{3.20}$$

Although, the Hampel minimum function is a non-convex function, which leads to a redescending influence function ψ, it has some interesting properties, which have made its performance very successful. It should be noted, although outliers are a serious problem and must be formulated in estimation process, Gaussian noise is also present. Therefore, the minimum function must be able to handle outliers and Gaussian noise simultaneously. In this case, it makes sense to design a function that resembles a least squares norm for small and Guassian errors, but at the same time rejects the extreme outliers entirely, which implies that the weight function vanishes outside some central regions. This concept of handling the combination of well-behaved noise and outliers simultaneously is partly realized in the Hampel function. The first part of the function resembles a Gaussian function to count for the normally distributed noise errors and yield a unique minimum. The second part of the function is equivalent to a Laplacian ($L_1 - norm$) minimization function in order to reduce the effect of the errors with moderate size and in the same time keep the function convex. Therefore it guarantees convergence for linear systems or in the presence of good approximation values for the non-linear systems. The third part of the function still reduces the effect of the gross errors but is not convex. That means, no convergence is guaranteed if the first initial guess starts at this region. The last part of the function designed to eliminate the effects of the outliers is completed by putting an upper bound. The figure (3.3) illustrate the graphical representation of the Hampel- minimum (3.3-a), influence (3.3-b) and weight (3.3-c) functions .



a b c

Figure 3.3: Hampel three-part M-estimator functions: a) Minimum function $\rho(x)$, b) Influence function $\psi(x)$, c) weight function $w(x)$

Note that the weights cannot be computed without an estimate of the residuals, which in turn requires knowledge of the solution. In addition, an estimation of the scale s, of the non-corrupted data is required. That is equivalent to robust estimation of the standard deviation σ, of the residual errors. If a good estimate of the standard deviation of the errors of non-outlier data can be made, then data points which lie beyond a certain number of standard deviation from the center can be classified as outliers. The standard deviation of the error σ, is either known a priori or can be found as a maximum likelihood estimate using MAD function which is formulated in equation (3.23) and is discribed in next section. Huber (1981) suggests an iterative computation scheme to minimize the error function. The minimization can be applied either by modifying the residuals or weights. In the following we use the modified weights method in the proposed estimation algorithm in which the weights are held constant at values equal to those found at the last iteration, whilst the set of parameters is estimated. This method follows the principles of the standard iterative least squares (LS) estimation process, where the solution is obtained based on the equation (3.7), with the exception that the corresponding weight matrix P, is determined by the equation (3.17), depending on the current residuals v_i. Huber proves that if these iterations are repeated a local (possibly global) minimum of the cost function (see equation 3.13), is reached. The algorithm is presented in section (3.4), with some modification with respect to its original procedure in the computation of the error term where the redundancy numbers are used in order to take the geometry of the observations into the account . However, in order to guarantee convergence and at the same time eliminating the effect of the outliers, even the iterated least squares approach suggested by Huber is not suitable unless there is a good initial value for the parameters or if there are a few gross outliers which are easily identified. This is our motivation to develop a two-stage estimation method discussed in section (3.4).

3.3.2 Random Sampling

The random sampling principle is based on the assumption that a subset of randomly sampled data points from a total set of points often provides a good estimate of the characteristics of the data set. Fischler and Bolles (1981), with their *random sample consensus (RANSAC)* algorithm were amongst the earliest to draw the value of such methods to the attention of computer vision researchers. It was a few years later that a similar robust parameters estimator called *least median squares (LMS)* was developed in the field of statistics by (Rousseeuw & Leroy 1987). The algorithms differ slightly in finding the global minimum for the associated cost function. The idea of random sampling has been applied in different forms and is reported in different applications (Bolles & Fischler 1981, Kumar & Hanson 1990, Roth & Levine 1990, Schunck 1990, Sinha & Schunck 1989, Torr & Murray 1993, Torr & Zisserman 1997, Gülch et al. 1998). Given that a large proportion of the data may be useless, the approach is opposite to the conventional least squares techniques. Rather than using as much data as is possible to obtain an initial solution and then attempting to identify outliers, a subset of data is used, as small as possible, to estimate the unknown parameters. For example, three points for estimating the initial parameters of a plane in 3D object space. In addition, in contrary to the standard LS estimation method, it theoretically has the maximum breakdown points of more than 50%. The random sample algorithm handles outliers by computing an estimate for model parameters from a consistent subset of inconsistent data points. The algorithm randomly selects a subset s, of the minimum number of points required to fit a model from the set of data points n. For each random sample of minimum size u, the initial model parameters are estimated and the cost function evaluated for each model. This process is repeated enough times on different subsets s, to ensure that there is a e.g. 95% chance that one of the subset will contain only good data points. The best solution is the one that minimizes the associated cost function. Minimizing the cost function in the RANSAC method is equivalent to the solution that maximizes the number of inlier points or the points whose error measure is below a threshold. While in the case of a LMS estimator it is equivalent to the solution that gives the minimum of the median of the square errors. Once outliers are detected, the set of points identified as inliers may be combined to give a final solution.

Ideally every possible subset of the data points would be considered, but in practice this is computationally infeasible. Fischler and Bolles (1981) and Rousseeuw and Leory (1987) proposed slightly different means of calculation for the required number of samples. But both give broadly similar numbers, we follow the latter suggestion. The number m of samples is chosen sufficiently high to give a probability p in excess of 95% that a good subset is selected. The expression for this probability p is:

$$m \geq \frac{\log(1-p)}{\log(1-(1-\varepsilon)^u)} \tag{3.21}$$

where ε is the fraction of contaminated data, and u the number of minimum data points in each sample. Table (3.1) gives some sample values of the number m of subsamples required to ensure $p \geq 0.95\%$ for given u and ε. It can been seen that far being computationally prohibitive, the algorithm may require less repeat than there are outliers, as it is not directly linked to the number but only the proportion of outliers. It can also be seen that the smaller the number of data points needed to estimate the model parameters, the fewer samples are required for a given level of confidence. If the fraction of data that is contaminated is unknown, as it is usual, an educated worst case estimate of the level of contamination must be made in order to determine the number of samples to be taken, this can be updated as larger consistent sets are found. In practice, however, it is recommended to take more subsamples than are needed, as some of the subsets might lead to degenerate solutions. For example when fitting a plane into the 3 collinear points in object space.

In order to detect and remove outliers from the data, some knowledge of the standard deviation σ of the error is required. In practice, outliers are discriminated from inliers based on the following equation:

$$i \in \begin{cases} inliers & if \quad d_i \leq t = c\sigma \\ outliers & otherwise \end{cases} \tag{3.22}$$

where t is a user defined disparity threshold, and d is the error measure, i.e., in the case of finding the best planar fit into the mesh of 3D points, d_i is the orthogonal distance between point i, and the estimated plane. Often the value of σ is unknown, in which case it must be estimated from the data. If there are no outliers in the data then σ can be estimated directly as the standard deviation of the residuals of a least square minimization process. If there are outliers and they are in the minority, a first estimate of the σ can be derived from the equation (3.23),

u	$\varepsilon = 5\%$	$\varepsilon = 10\%$	$\varepsilon = 20\%$	$\varepsilon = 25\%$	$\varepsilon = 30\%$	$\varepsilon = 40\%$	$\varepsilon = 50\%$
2	2	2	3	4	5	7	11
3	2	3	5	6	8	13	23
4	2	3	6	8	11	22	47
5	3	4	8	12	17	38	95
6	3	4	10	16	24	63	191
7	3	5	13	21	35	106	382
8	3	6	17	29	51	177	766

Table 3.1: The number m of subsets required to ensure $p \geq 95\%$ for given u and ε, where p is the probability that all the data points selected in one subset are non-outliers (courtesy of Torr and Murray 1993).

based on the median of the absolute value of the errors of the chosen parameter fit. This function is known as *median absolute deviation* (MAD) function in the literature (Huber 1981, Rousseeuw & Leroy 1987).

$$\sigma = \frac{med|d_i|}{\Phi^{-1}(0.75)} \tag{3.23}$$

It is known that equation (3.23) is an asymptotically consistent estimator of σ when d_i follow a Gaussian distribution $N(0, \sigma^2)$, and where Φ is the cumulative distribution function for the Gaussian probability density function. It was shown empirically (Rousseeuw & Leroy 1987), that when $n \approx 2u$ the correction factor of $(1 + \frac{5}{n-u})$ improves the estimate of the standard deviation. Noting that $\Phi^{-1}(0.75) = 0.6745$ the estimate of σ is then

$$\sigma = (1 + \frac{5}{n-u})\frac{med|d_i|}{0.6745} \tag{3.24}$$

In our proposed synthesis robust estimator algorithm in section (3.4), we suggested to use the LMS algorithm if there is no a priori knowledge about the σ, in order to obtain the estimate of the median and compute the first estimate of the standard deviation. Thus the outlier data points are classified and consequently the initial values of the unknown model parameters are estimated. The analysis of the test results obtained by fitting the 3D plane polygons over the existing DSM showed that random sampling techniques can provide a good initial guess of the plane parameters at a solution, but this solution can bear improvement as usually not all the outliers will be detected. Earlier it was noted that iterative estimation of the M-estimators is only successful if the starting estimate was good. By using the output of the random sampling rather than linear regression as the starting estimate for M-estimation, here using an iterative Huber algorithm, a further improvement can be made. This is our key conclusion to introduce a new two-stage robust parameters estimation which will be elaborated in more details in section (3.4). The results are also indicated that RANSAC is superior with respect to LMS, firstly when the standard deviation σ of the error term was known and secondly, when there are more than 50% outliers in the original data.

3.3.3 Clustering

The concept of clustering is used in a variety of related methods for parameter estimation (Sester & Förstner 1989, Roth & Levine 1990, Schickler 1992, Gülch et al. 1998), and it follows the principle of maximum likelihood estimation. In fact, a classical example is the *Hough transform (HT)* technique, which has long history of valuable service to computer vision for detection of simple shapes such as straight lines, or circles in the image. It uses a parametric description of simple geometric primitives in order to reduce the computational complexity of their search in the data set. This is realized in an accumulator array by partitioning the parameter space into cells where every dimension of this space is quantized into specified intervals. The performance and accuracy of the method depend of course on the quantization interval of the space parameters, and on the size of the accumulator array. Each data point adds a vote to every cell of parameter space (corresponding element of

the accumulator array incremented by 1), whose combination of parameters could have produced a version of the interested geometric primitive. When all the data points have been processed, the cells in the parameter space which are local maxima or which have a number of votes greater than a given threshold are marked as representing possible solutions since they are well described by the primitive whose parameters is associated with those cells. The clustering technique is recommended for application with few unknowns, high percentage of outliers and a high redundancy in order to support the solution (Gülch et al. 1998). However, it runs into problems when the dimension of the parameter space is high, because its space requirements is exponential in the dimensionality and the computational expense rises exponentially with the dimension of the parameter space. Alternative solutions of a coarse quantization decrease the accuracy and reliability of the method.

3.3.4 Case Deletion Diagnostics

The principles of case deletion methods are based on influence measures (Chaterjee & Hadi 1988). The basic concept is simple. Small perturbations are introduced into some aspects of the model formulation and an assessment is made of how much these change the outcome of the analysis. The important issues are the determination of the type of perturbation scheme, the particular aspect of the analysis to monitor and the method of assessment. Several different measures of influence have been proposed within statistical literature, mainly based on the eigenvalues and eigenvectors of the covariance matrix of the estimated parameters. Critchley (1985) suggested the use of eigen-perturbation to arrive at influence functions assessing the first or higher order effect on the principal eigenvalues and eigenvectors. Shapiro and Bardy (1993) proposed an influence measure that monitors the effects of the deletion on the minimum eigenvalue. Torr and Murray (1993) reported an interesting method for the case of orthogonal regression based on eigenvector perturbation theory. They define a parameter, so-called *leverage factor*, which gives a measure of the influence of each point and is large for outliers even when the residual is small. To remove outliers from the estimation model, a point with maximum influence is deleted in each iteration and the regression process is recomputed. This procedure is repeated until the termination criteria is met. The disadvantage of the case deletion schemes is that they require a fairly good estimate of the standard deviation σ of the error term in the data set. In addition, they rapidly breakdown after 35% outliers, and only provide inaccurate results below that.

3.3.5 Minimum Description Length

The method of *minimum description length (MDL)* offers a different concept as a robust parameters estimator for detection of outliers (Rissanen 1987). It has its root in information theory and is mainly a tool for describing the data, thus obtaining information for comparing the hypothesis parametric models with varying complexity. If several models are suggested, then the model giving the shortest description length should be chosen. This ability of comparing different parametric models can be used for different applications depending on the formulation of the problem (Förstner 1989). Axelsson (1996) studied the application of MDL criteria in different projects related to image orientation procedures, photogrammetric measurement processes and correspondence problem of multiple images, in order to localize and eliminate the erroneous observations in estimation processes. The thresholding problem for removing the outliers is not present since a minimum of the description length is found. However, if the data cannot be modeled in a suitable manner, the description lengths of the best model parameters will be higher than the one of unmodeled data. It has, theoretically, a high breakdown point of more than 50%, but there is no simple analytical solution. Instead a numerical search procedure is applied. That means the solution is not optimal in minimizing the error term for the remaining inlier points, similar to most robust estimators with high breakdown points. Therefore, a least squares type estimator is utilized at the final process.

3.4 Synthesis Robust Estimators

Although the robust estimation techniques are far superior to non-robust methods when the original data are contaminated by outliers, they are still of course imperfect. The M-estimator, which is more satisfying from a statistical standpoint, is only successful if the starting estimate is good. It rapidly breaks down when the amount of outliers exceeds 35% of the data set and only provides inaccurate results after this limit. On the contrary, the random sampling techniques have high breakdown points and can provide a good first guess at a solution, but this solution still needs improvement, as usually not all the outliers are detected. These observations have allowed us to propose an empirically optimal two-stage robust parameter estimation process. By combining the

high breakdown points of the random sampling methods and the high efficiency of the M-estimator methods, here the iterative Huber algorithm based on Hampel three-part redescending minimization function (see section 3.3.1), a further improvement can be made. The first stage flushes and detects the outliers and estimates the best initial model parameters based on the remaining inliers data using a random sampling type estimator such as RANSAC. The estimated parameter values along with the estimated error variance $\hat{\sigma}$ are introduced into the iterative re-weighting M-estimator algorithm as initial values to compute the final model parameters. Figure (3.4) illustrates the improvement afforded using the proposed synthesis robust parameter estimation method in a 3D regression problem.

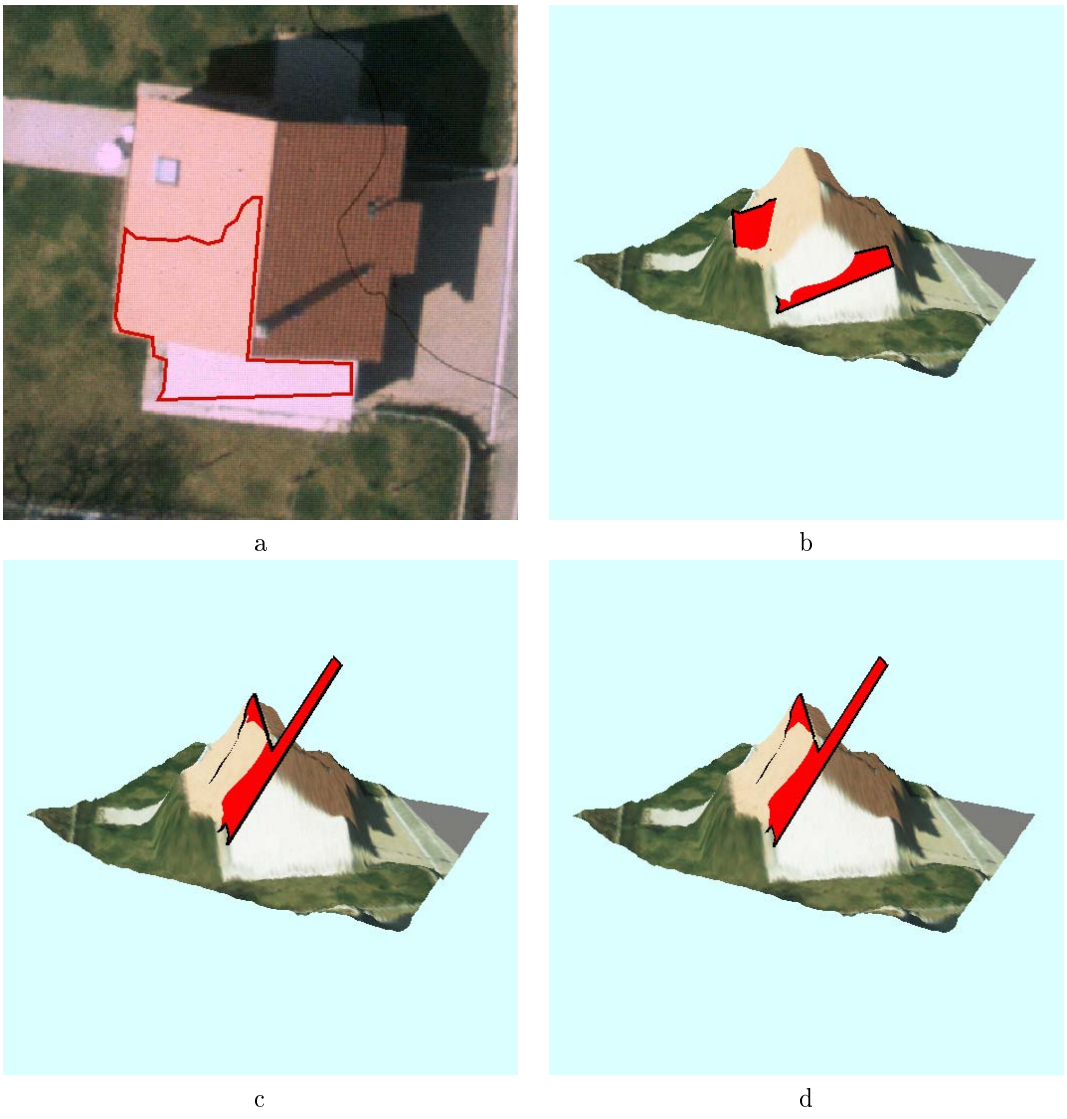


Figure 3.4: Estimated 3D plane-roof polygon of a building roof overlaid on the 3D perspective view of the correponding DSM: a) 2D plane-roof region overlaid on corresponding roof structure, b) corresponding 3D plane-roof polygon back projected into the object space based on a standard LS estimation process, c) corresponding 3D plane-roof polygon back projected into the object space based on the RANSAC process, d) corresponding 3D plane-roof polygon back projected into the object space based on the synthesis robust estimation techniques

During the reconstruction process the extracted 2D plane-roof region in image space is back projected into the 3D object space (see chapter 5). This is performed by fitting the corresponding 3D plane-roof polygon over the existing DSM, in order to determine the best parameters $x(a, b, d)$ of the 3D plane, which is defined explicitly by equation (3.25).

$$z = ax + by + d \tag{3.25}$$

The best optimal solution is achieved, when the orthogonal geometric distances l_i between all the given inlier points i and the estimated 3D plane are minimum.

$$l_i = \frac{z_i - (ax_i + by_i + d)}{\sqrt{a^2 + b^2 + 1}} \tag{3.26}$$

This is equivalent to the standard least squares estimation criteria when the errors are normally distributed and the extreme outliers are eliminated from the data set. Figure (3.4-a) illustrates one side of a plane-roof structure of a residential building, which is detected and segmented based on a least squares planar fit region growing algorithm (chapter 4). Assuming that the low contrast white area in the lower part of the building causes the segmentation algorithm to grow over the bounding edge of the roof, and therefore incorrectly extract this area as part of the 2D plane-roof region. The data points belonging to this part of the roof appear as outliers during back projection of the extracted 2D plane-roof region into the 3D object space and should be detected and eliminated from the 3D regression process. The figures (3.4-b), (3.4-c), and (3.4-d) illustrate the 3D perspective views of the estimated 3D planes obtained by an ordinary LS estimation, the RANSAC procedure, and our proposed method repectively. Figure (3.4-b) indicates graphically the failure of the LS procedure in estimating the parameters of the 3D plane, which is forced by the contaminated data points, while two other approaches correctly recovered the parameters of the 3D plane.

The angle α, between the normal vector of the reference 3D plane measured in a stereo model and the normal vectors of the estimated 3D planes based on the above approaches are computed and tabulated in table (3.2) for the comparison. The numerical results also show the failure of the LS method in recovering the parameters of the plane in the presence of the outliers. Instead of approximating the building roof, the estimated 3D plane has cut off the building (figure 3.4-b). In addition, the small improvement of the result based on the synthesis robust method with respect to RANSAC is monitored. Although the improvement appears small, it has a significant effect in an automated vision process, as it is discussed in chapter 5. Note that the estimated value of the $\hat{\sigma}$ in the RANSAC procedure indicates a better fit. This is expected because the standard deviation in RANSAC method is only computed over the inlier data points.

Methods	α	$\hat{\sigma}$
Least squares	$41°27'$	1.0
RANSAC	$18°10'$	0.19
Synthesis method	$17°53'$	0.21

Table 3.2: Comparison of the 3D reference plane with the corresponding 3D plane-roof polygons computed based on different estimation process.

The relatively large values of α, even utilizing the robust parameter estimation techniques can be explained by the fact that estimated 3D planes are computed based on the DSM, which is itself an approximation of the original building. Therefore, in this case, the value of the angle α is a good indication of the quality of the existing DSM.

The procedure of the proposed method is summarized as a pseudo-code for the 3D plane-roof polygon regression and it can be extended and applied to different estimation problems by utilizing the appropriate mathematical functions.

- Stage I: Estimating the first guess of the model parameters based on the random sampling techniques.

  1. Repeat for m sampling as determined in equation (3.21):
     (a) Select a random sample of the minimum number of data points, i.e. 3 points, to compute an estimate of the model parameters, i.e. $\hat{x}(a, b, d)$
     (b) Calculate the distance measure l_i for each data point given $\hat{x}$, i.e. using equation (3.26)
     (c) Apply RANSAC, if there is a priori knowledge upon the standard deviation σ of the error term:

i. Compute the normalized error measure t_i for each data point based on the following equation (Baarda 1968)

$$t_i = \frac{l_i}{\sigma\sqrt{r_i}} \tag{3.27}$$

where r_i is called redundancy number and is the (i,i) element of the redundancy matrix $\mathbf{R}$, defined using the equation (3.28)

$$r_i = R_{ii} = (\mathbf{I} - \mathbf{A}(\mathbf{A^T P A})^{-1}\mathbf{A^T P})_{ii}. \tag{3.28}$$

Recall that $\mathbf{A}$ and $\mathbf{P}$ are the jacobian and weight matrices of a Gauss-Markov estimation model (see equation 3.5) respectively and $\mathbf{I}$ is the unit matrix. In this manner, the local geometry of the data points in the design and therefore the effects of the gross errors onto the distance measures are taken into account.

ii. Calculate the number of inliers consistent with estimated parameters $\hat{x}$ based on the following statistical test:

$$data\ point\ i \in \begin{cases} inliers & if \quad t_i \leq t \\ outliers & otherwise \end{cases} \tag{3.29}$$

where t is a user defined disparity threshold (e.g. $t = 1.96$).

(d) Else, apply LMS:

i. Define the median absolute error for the selected sample $med|l_i|$.

ii. Compute the normalized error measure t_i for each data point based on the equation (3.27), where the standard deviation of the error term σ is estimated based on the MAD function (3.23).

iii. Calculate the number of inliers consistent with estimated parameters $\hat{x}$ based on the equation (3.29).

2. Select the best solution:

(a) If RANSAC, we obtain the solution with the maximum number of inliers. In the case of ties select that solution which has the lowest standard deviation of inliers residuals.

(b) If LMS, the solution which gives the minimum median error is obtained. In the case of ties select that solution which has the lowest standard deviation of inliers residuals.

3. Re-estimate the model parameters $\hat{x}$ and the standard deviation of the error $\hat{\sigma}$ using all the data that has been identified as consistent (inliers) and passes these parameters into the second stage of the estimation process as the initial values of the model parameters.

- Stage II: Estimating the final model parameters using an iterative re-weighting M-estimator.

1. Repeat until the termination criteria met or when the maximum number of iteration is reached:

(a) Calculate the distance measure l_i for each data point using equation (3.26) based on the current estimated values of the model parameters $\hat{x}$.

(b) Compute the normalized error measure t_i for each data point based on the equation (3.27), and the current standard deviation of the error term $\hat{\sigma}$.

(c) Calculate corresponding weight w_i of each data point based on the Hample three-part weight function (3.20). Note that the required modification to the computed weights w_i and σ as it is proposed by (Huber 1981, pp. 183) should be taken into account.

(d) estimate corrections of the model parameters, i.e. Δx, compute the new model parameters $x^{n+1} = x^n + \Delta x$, and the standard deviation $\hat{\sigma}$ based on a weighted least squares solution.

2. Save the last estimated parameters $\hat{x}$ as the final model parameters, i.e. the final parameters of the 3D plane-roof polygon.

Chapter 4

From Pixels to Geometric Primitives

4.1 Introduction

Object recognition is one of the hardest problem in automated vision processes. Although the problem is addressed by a large number of researchers and projects (Tou & Gonzalez 1974, Lowe 1985, Canny 1986, Fua & Hanson 1987, Mohan & Nevatia 1989, Strat & Fischler 1991, Haala & Vosselman 1992, Haralick & Shapiro 1992, Gonzalez & Woods 1993, Brunn et al. 1996, Sagerer, Kummert & Socher 1996, Fritsch & Ameri 1998, Salesh & Sowmya 1998), there is not a unique baseline methodology or a general paradigm to solve this complex task. Based on the fact that simple geometric primitives are an important part of the human visual perception, this task is normally initialized with an image segmentation process, where the qualitative knowledge stored in the raw image data is transferred into quantitative symbolic description and more abstract form of the basic geometric elements. Image segmentation is frequently a data driven process. It can be based on homogeneities, namely, homogeneous regions that are detected in the image, or alternatively, discontinuities that can be used for the detection of edge primitives, which is assumed to correspond to the contours of the real objects. This chapter deals with the recognition of 2D planar surfaces in the aerial image, in this study called *2D plane-roof regions*, which possess meaningful correspondence to object surfaces in 3D scene, here *3D plane-roof polygons* of the building roof structures. The recognition performs in three subsequent segmentation processes. First, the image is partitioned into the regions of interest based on height discontinuities and their size. Those areas, which have a high expectancy of being and representing the individual building are extracted. This process is performed by a morphological top-hat transformation of the corresponding DSM (Weidner & Förstner 1995). Furthermore, a coarse segmentation of the image is carried out based on the geometric characteristics of surfaces (Besl & Jain 1988). The *mean* and *Gaussian curvatures* are used to extract flat pixel surfaces type within every region of interest. Finally, the fine segmentation of the image is performed. The extracted 4-connected flat pixels serve as the seed regions to a least squares planar fit region growing algorithm to partition the image surface into meaningful primitive plane-roof regions (Fritsch & Ameri 1998). The intermediate extracted symbolic 2D plane-roof regions are projected back into the object space, in order to reconstruct the corresponding plane-roof polygons in 3D space, which is the subject of the next chapter. Each part of the recognition process is described in detail in subsequent sections. The main emphasis is given to the geometric characteristics of digital surfaces (section 4.3), extraction of regions of interest (section 4.4), and an iterative region-based segmentation algorithm (section 4.5). All the proposed methods are implemented and the results are presented.

4.2 Why Region-based Segmentation

Recall from the introductory chapter, it is discussed that the whole framework of our method is based on a *data-driven* reconstruction of a *generic polyhedral-like* building model. This is motivated from the fact that building roofs are mostly an aggregation of planar surfaces. Reconstruction based on a generic object model means that the number, as well as the geometric form, and the position of the significant parts of the model have to be defined. In addition, the geometric and topological relationships between these primitives are also needed. Finding the logical relationships between these geometric primitives when a specific object model is not present, is a complex problem, and its complexity is in a reciprocal-like relation with the geometrical level of the incorporated geometric primitives. That means, hypothesis model generation of a generic object based on point or line primitives is more complex than a polygonal-based approach.

The following simple example is deliberated to illustrate the concept. Assume an object, e.g. a roof structure of a building consists of n polygons $P_i, i = 1, \cdots, n$, where each polygon P_i, is defined by m bounding edges $E_j, j = 1, \cdots, m$, or alternatively l vertices $V_k, k = 1, \cdots, l$. Let us consider the particular case of a convex polygon $m = l$. In fact, the mutual relationships between these geometric elements in the absence of any external or supporting knowledge can be explained by *combinatorial mathematics*, which define the total numbers of

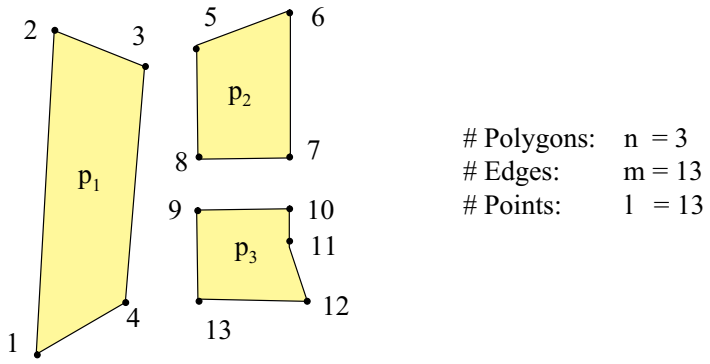


Figure 4.1: Object structure represented by its geometric primitives

possible combinations of s elements (selections) from a set of q distinct elements and is determined by the following equation:

$$C(q,s) = \frac{q!}{(q-s)! \cdot s!} \tag{4.1}$$

where q is the number of individual elements of the set, and s indicates the number of elements contributed into the combination. Let us further consider a particular case of mutual relationships, $s = 2$. The total number of possible combinations between the geometric primitives of the object presented in figure (4.1) for three different cases based on only (I) point primitives e.g., $q = 13$, (II) line primitives e.g., $q = 13$, and (III) polygonal primitives e.g., $q = 3$, is tabulated in table (4.1).

	I	II	III
S	2	2	2
C	78	78	3

Table 4.1: Maximum possible number of combination between s randomly selected elements from a set of q distinct elements

The computed values indicate that the maximum numbers of possible combinations between the higher level geometric primitives i.e. polygons, are significantly smaller than those in lower level i.e. points or edges. Moreover, since the objects, which we are interested in describing (roof structures of the building) are mainly made up of plane surfaces, therefore, this is a natural choice to partition pixels in the image into the regions that possess meaningful correspondence to object surfaces in a 3D scene. In addition, regional information provides helpful clues in automated 3D image analysis, they do provide many descriptions such as area, surface normal, average grey level, etc., which are not derivable from edges or line segments. They also provide topological information, in case that the polygon adjacency relationships (PAR) is computed, that in turn give us the ability to make queries such as which region is a *contained-in* region, and so on. In most related works for building reconstruction, edges or lines are extracted from the image(s) as the basic image features for further analysis. Often, these lines are the only sources of information for solving the task, thus disregarding the original images and their regional information. These image features often have no relation to each other because correctly linked edges could not be extracted from real images and that they are unreliable close to intersections. In edge-based approaches, it is difficult to measure the edge error directly against the original image data because an explicit edge description is not present in the original data. Region-based algorithms may potentially have an advantage over edge-based algorithms because it is possible to check final image interpretation against the original data at every image pixel via simple image subtraction.

In the proposed method, a region-based segmentation algorithm based on the geometry of the surfaces initializes the recognition process. In fact, in real-world images, the structures of the imaged object cannot be detected solely on the basis of their registered photometry because of the presence of noise, occlusion and various photometric anomalies. Therefore, segmentation methods based on purely local statistical criteria are tied to errors

(Fua & Leclerc 1990). To supplement the weak and noisy local information of the images, geometric information is also incorporated into the recognition process. In addition, it is recognized that discontinuities should be represented by line segments, which can also be detected with sub-pixel accuracy and thus give a high quality result. The region-based segmentation algorithms may miss the relevant boundary information, and are generally unable to trace fine detail and linear elements. They usually tend to produce regions of which the shape reflects more the search strategy used than the true shape of the regions (Lemmens 1996). In consequence, an automated 3D image analysis should incorporate the descriptions of point, line, and region segments to admit a compact transfer of most of the information content in the image to higher level processes. This is the key issue and the strength of our proposed method. It enters into the high-level quantitative domain of the recognition process –extraction of regional information– in order to reduce the complexity of the problem in the very early stage of the whole chain of a generic model-based reconstruction process, and integrates the qualitative geometric primitives –point and edge information–in the high-level model-oriented process during the hypothesis verification process, which is discussed in chapter 6. This is in contrary to the most of the reported methods, which initialize their recognition process from the low-level geometric primitives, and struggling with complex search strategies in the higher level processes.

4.3 Geometric Characteristics of Surfaces

Geometrical proximity plays an important role in image segmentation. Pixels lying in the same neighborhood tend to have similar statistical properties and belong in the same image region. Thus an image segmentation algorithm must incorporate, if possible, both proximity and homogeneity to produce connected image regions (Pitas 1993). It has been shown by Besl (1988), that differential geometric concepts for visible-invariant descriptions of continuous surfaces are applicable to digital surfaces even in the presence of quantization and measurement noise. That is each point on a continuous or digital surface can be characterized by the spatial properties of other points on the surface in small neighborhoods surrounding the given point. The key difference is that the neighborhood of a point consists of an uncountable infinite number of points in the continuous surfaces whereas small finite numbers of points form the neighborhood of a point in a digital surface. In other words, the digital images are sampled graph surfaces.

In general, an explicit form of expressing a surface is the graph of a function of two variables $f(x, y)$. In the context of computer vision, grey level surfaces in intensity images, and depth surfaces in range images conform to this common representation and can be analyzed in this way. If x and y denote spatial coordinates, then the value of f is proportional to the brightness of the image or is the distance from the camera origin to the surface at the point (x, y) respectively. The general surface S in explicit parametric form is defined as :

$$S = \{(x, y, z) : x = d(u, v), y = e(u, v), z = f(u, v) \mid (u, v) \in D \subseteq R^2\}. \tag{4.2}$$

In this study, we consider only smooth surfaces, where all three parametric functions have continuous second partial derivatives. In general, an intensity image or a range image may have several smooth surfaces separated by points of discontinuity, e.g. step edges or orientation edges. In this particular case, we may rewrite equation (4.2) in a less general form which is equivalent to the graph surface form as:

$$S = \{(x, y, z) : x = u, y = v, z = f(u, v) \mid (u, v) \in D \subseteq R^2\}. \tag{4.3}$$

The geometry of such a surface depends on two classical quadratic differential forms called the *first* and *second fundamental forms*. Complete knowledge of either of these forms at every surface point provides an analysis and classification of smooth surface shape. For a given surface or surface patch $S(u, v)$, the first fundamental form $I(u, v, du, dv)$ is defined as:

$$\begin{aligned} I(u, v, du, dv) &= dS \cdot dS = Edu^2 + 2Fdudv + Gdv^2 \\ &= [du \quad dv] \begin{bmatrix} j_{11} & j_{12} \\ j_{21} & j_{22} \end{bmatrix} [du \quad dv] \end{aligned} \tag{4.4}$$

where the elements of the symmetric matrix **J** are

$$j_{ii} = E = S_u \cdot S_u \quad , \quad j_{22} = G = S_v \cdot S_v \quad , \quad j_{12} = j_{21} = F = S_u \cdot S_v \tag{4.5}$$

and the subscripts denote the first partial derivatives of the surface defined as:

$$S_u(u,v) = \frac{\partial S}{\partial u} \quad , \quad S_v(u,v) = \frac{\partial S}{\partial v}. \tag{4.6}$$

The $S_u(u,v)$ and $S_v(u,v)$ define the tangent vectors to the surface at the point (u,v). They form the basis of the tangent plane which touches the surface at point (u,v). The matrix J is the first fundamental form matrix, the *metric* or the *metric tensor* of the surface. The first fundamental form $I(u,v,du,dv)$ is a measure of a small amount of movement on the surface at a point (u,v) for a given small movement in the parameter plane (du,dv). It is invariant to translations and rotations of the surface in 3D space, and thus is an *intrinsic* property of the surface, that means it depends only on the surface itself, not on how the surface is embedded in 3D space. The second fundamental form $II(u,v,du,dv)$, defined as:

$$\begin{aligned} II(u,v,du,dv) &= dS \cdot d\vec{n} = Ldu^2 + 2Mdudv + Ndv^2 \\ &= [du \quad dv] \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} [du \quad dv] \end{aligned} \tag{4.7}$$

where the elements of the symmetric matrix $\mathbf{B}$ are

$$b_{ii} = L = S_{uu} \cdot \vec{n} \quad , \quad b_{22} = N = S_{vv} \cdot \vec{n} \quad , \quad b_{12} = b_{21} = M = S_{uv} \cdot \vec{n} \tag{4.8}$$

and $\vec{n}(u,v)$ is unit normal vector and defined as:

$$\vec{n} = \frac{S_u \times S_v}{||S_u \times S_v||}. \tag{4.9}$$

The double subscripts denote second partial derivatives of the surface and are defined as:

$$S_{uu}(u,v) = \frac{\partial^2 S}{\partial u^2} \quad , \quad S_{vv}(u,v) = \frac{\partial^2 S}{\partial v^2} \quad , \quad S_{uv}(u,v) = \frac{\partial^2 S}{\partial u \partial v} \tag{4.10}$$

$II(u,v,du,dv)$ measures the correlation between the change in the normal vector $d\vec{n}$, and the change of surface position dS at a surface point (u,v) as a function of a small movement (du,dv) in the parameter plane. Therefore, it depends on the position of the surface in 3D space and thus is an *extrinsic* property of the surface. The differential normal vector $d\vec{n}$ always lies in the tangent plane. The ratio of $II(u,v,du,dv)/I(u,v,du,dv)$ is the *normal curvature function* K_{normal}, which varies as a function of direction of the differential vector (du,dv). If $d\vec{n}$ and dS are aligned for a particular direction of (du,dv) then that direction is a *principal direction* of the surface at that surface point. The extrema of the normal curvature function occur at that point and are called the *principal curvatures*, k_1 the *maximum*, and k_2 the *minimum*.

4.3.1 Mean and Gaussian Curvatures

It is established that a surface may be characterized by six functions E, F, G, L, M and N derived by equations (4.5, 4.8). It has also shown that the information from these functions can be reduced into two curvature functions k_1 and k_2. There are several other combinations of these functions that yield more easily interpretable surface shape characteristics. Particularly, there are two curvature functions, *mean* H and *Gaussian* K curvatures that combine the information in these six functions.

These two curvature functions do not, in general, contain all the 3D shape information contained in the six E, F, G, L, M, N functions and some of the information has been lost, but they do contain a substantial amount of information. They are direction independent quantities and can be computed by combining the two direction-dependent principal curvatures k_1 and k_2 as follows:

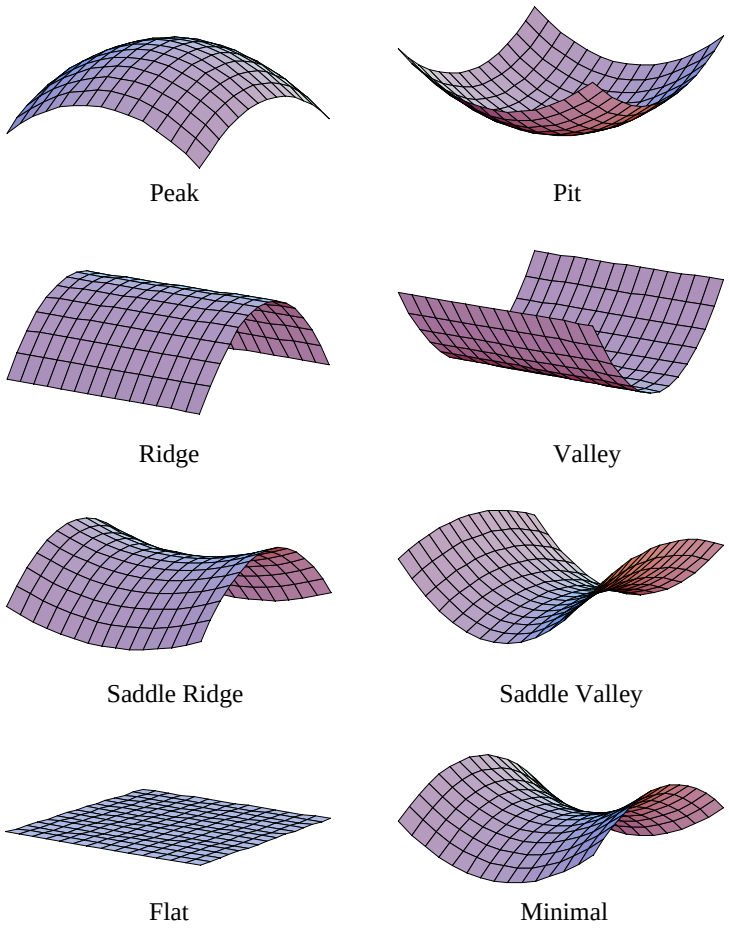


Figure 4.2: Eight fundamental surface types defined by mean and Gaussian curvatures signs (Courtesy of Besl 1988)

$$K = k_1 \cdot k_2 \tag{4.11}$$

$$H = \frac{k_1 + k_2}{2}. \tag{4.12}$$

In fact, k_1 and k_2 are roots of the quadratic equation:

$$k^2 - 2Hk + K = 0. \tag{4.13}$$

Therefore, if K and H are known at each point, it is straight forward to analytically determine the two principal curvatures:

$$k_{1,2} = H \pm \sqrt{H^2 - K}. \tag{4.14}$$

The mean and Gaussian curvatures are important quantities in computer vision because they provide a common method of specifying eight basic types of surfaces surrounding any point on a smooth surface (figure 4.2), which is discussed in more detail in section (4.5.2). The method based on k_1 and k_2 is less common as they are dependent on directions of maximal and minimal normal curvatures at each point, whereas mean and Gaussian curvature values are direction-free quantities. The pair k_1, k_2 contain the same surface curvature information as the pair H, K but in a different form. There can be advantages and disadvantages working with either pair depending on the application. For visible-invariant pixel labeling purposes, the sign of mean and Gaussian

curvatures can be computed most easily yielding the coarse classification of surface types in image data (Besl & Jain 1988, Fritsch & Ameri 1998).

4.4 Regions of Interest

The overall aim of our approach in this section is to partition an image into regions, which have potential to be the buildings. This primary segmentation has been done in order to, 1) reduce the dimensionality of search space and consequently reducing the computational time required in subsequent processes, and 2) to be a step closer to our ultimate aim. In the absence of GIS information or ground plan of the buildings, the corresponding DSM is used as an initial source for building detection. However, the proposed method is general enough to integrate the contribution of existing data in this step, if any. A DSM is a geometric description or reconstruction of the physical sensed surface and can be considered as a noisy sample of the visible surface. It provides information about the objects which have been characterized by their relative heights respect to their surrounding, e.g. buildings, trees. It can be generated directly using a laser scanner sensor (Lohr & Eibert 1995), or using stereo images (Ackermann & Krzystek 1991, Schenk & Toth 1991). The quality of the DSM is an important issue in our reconstruction method. Indeed, the results of the mid-level processes in extraction and structuring 3D primitives is highly dependent on the quality of the utilized DSM, which in the worst case leads to partially or completely wrong descriptions of the buildings. This concept is elaborated in more details in the following chapters. Figure (4.3) shows a 3D perspective view of an image wrapped over a corresponding DSM. The figure illustrates the presence of the standing objects such as buildings and trees –as it is expected– in DSM.

Figure 4.3: 3D perspective view of an image wrapped over corresponding DSM

Several methods for detecting building candidates, *regions of interest*, from the DSM have been introduced so far; 1) The most simple and accurate one is subtraction of DSM from existing DTM (Digital Terrain Model) [1], but in most cases a DTM with sufficient resolution and accuracy is not available. 2) Extraction of 3D blobs or hight bins based on grouping the DSM heights into consecutive height ranges of a certain size (Baltsavias et al. 1995). 3) Applying morphological operators on the DSM to compute an approximation of terrain surface using grey opening (Weidner & Förstner 1995), or dual rank filter (Eckstein & Munkelt 1995). In this study we have used the grey opening approach proposed by Weidner and Förstner (1995), which is based on a square structuring element. Although the result is relatively satisfactory in the open areas, however, further investigations on

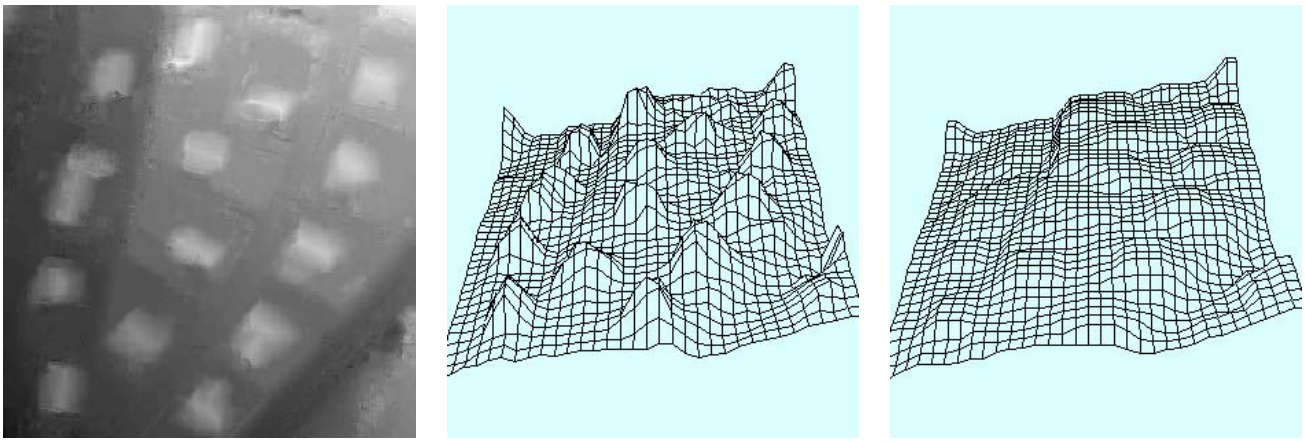
[1] A DTM is a geometric reconstruction of the terrain surface where buildings or other standing objects are considered outliers and are excluded.

the exploration of the potentials of the other approaches such as image classification using height data as a supporting channel (Walter 1999), or wavelet transformation (Fatemi Ghomi 1997, Wouwer 1998), specially in built-up areas, as alternative methods have to be carried out. Even a minor improvement of the results in this stage will significantly increase the quality and outcome of the subsequent processes (Ameri & Fritsch 1999).

Grey opening of a digital surface has a simple geometric interpretation (Gonzalez & Woods 1993). Suppose that we view a digital surface $z = S(x, y)$, in a 3D perspective (see figure 4.4-b), where x and y axes being the usual spatial coordinates and the third axis z being height. In this representation, a DSM appears as a discrete surface whose value at any point (x, y), is that of S at those coordinates, that is z. Let us now assume that we want to open S by a spherical structuring element k, and view this element as a rolling ball. The mechanisms of opening S by k may be interpreted geometrically as the process of pushing the ball against the underside of the surface, while at the same time rolling it so that the entire underside of the surface is traversed. The opening $S \circ k$ then is the surface of the highest points reached by any part of the sphere as it slides over the entire undersurface of S. That means, all the peaks that were narrow with respect to the diameter of the ball were reduced in size and sharpness. In practical applications, opening operations usually are applied in intensity based images to remove small –with respect to the size of the structuring element– light details, e. g. peaks, while leaving the overall grey levels and larger bright features relatively undisturbed. Figure (4.4-c) shows an approximation of a terrain surface so-called DTM, resulted from a *morphological opening* of DSM using a square structuring element k, of the size $w \times w$. Indeed, this is a *morphological grey scale erosion* of the DSM by k followed by a *dilation operation* with the same structuring element, which is defined as:.

$$S \circ k = (S \ominus k) \oplus k \tag{4.15}$$

Figure (4.4-a) shows the corresponding DSM as a grey value based image where the grey levels indicate the heights. The lighter areas present buildings and other standing objects like trees.



a b c

Figure 4.4: Morphological opening of DSM: a) grey value based DSM image, b) 3D perspective view of DSM, c) 3D perspective view of computed DTM

As described above, if the size of structuring element k will be selected in such a way that it contains these areas, then the initial grey scale erosion removes these areas, but it also darkens the image –extends the holes–. The subsequent dilation again increases the brightness –compensating the effect of erosion in extending the size of holes– of the image without reintroducing the details, i.e. buildings or trees, removed by erosion (see figure 4.4-c). The size of k will be chosen based on a priori knowledge about the minimum size of buildings in the scene. In fact, the extraction of regions of interest is equivalent to the generation of normalized DSM, which from an image processing point of view is simply a *morphological top-hat transformation* of the DSM and is defined by $h = S - (S \circ k)$, where S is the input DSM, and h is the height of the standing objects, followed by a thresholding process (equation 4.16), based on a priori knowledge about the minimum height of the buildings h_{min}.

$$h(x,y) = \begin{cases} h & if \quad h > h_{min} \\ \\ 0 & if \quad h < h_{min} \end{cases} \qquad (x,y) \in DSM \tag{4.16}$$

Within the initial extracted segments there are some other segments that do not present buildings, such as small groups of trees or small hills. Therefore the primary segmented image undergoes another thresholding process based on minimum size of buildings in the scene. The final segmentation, that is the extraction and labeling of the regions of interest based on minimum size threshold will be done in an efficient way during a standard connected component analysis. Only those regions with an area bigger than a pre-specified size threshold will be selected and labeled.

Figure (4.5-a) illustrates a perspective view of the extracted regions of interest in 3D object space. Some of the regions (partly) still do not present buildings or building parts. These are due to presence of features such as standing trees or cars next to a building, or group of trees whose size is bigger than a specified size threshold. These incorrectly classified regions or region parts will be filtered out in the follow up processes. An alternative solution is the analysis of texture pattern (Haralick, Shanmugam & Dinstein 1973, Nagao & Matsuyama 1980, Sali & Wolfson 1992, Lee & Schenk 1992, Lee & Schenk 1998, Wouwer 1998), within each region of interest, in order to detect and exclude objects that are standing adjacent to the buildings but are not a building or part of a building, such as trees. The intermediate result has shown that integration of this analysis at this stage possibly could overcome this problem, but still more investigation is needed. The transformation of the extracted regions of interest into the 2D image space is done based on collinearity equation (6.10). Figure (4.5-b) shows the result of this transformation.

a b

Figure 4.5: Extracted regions of interest: a) the 3D perspective of the extracted region wrapped over coresponding DSM, b) extracted region overlaid on the corresponding aerial image.

4.5 Iterative Region-based Segmentation

The objective of segmentation in this study is to partition an image into regions. In the previous section, we approached this problem by finding regions of interest based on height discontinuities and their size using mathematical morphology. In this section, we discuss a region-based segmentation technique based on an iterative region growing approach to partition each region of interest into the primitive planar regions, which are parts of building roofs in the real world. In order to implement region growing, we need a rule describing a growth mechanism and a rule checking the homogeneity of the regions after each growth step. The growth mechanism is simple, at each stage k and for each region $R_i^k, i = 1, \cdots, n$ we check if there is a pixel in the 8-neighborhood

of each pixel of the region border. Before assigning such a pixel P to a region R_i^k, we check if the region homogeneity:

$$H(R_i^k \cup \{P\}) = TRUE \tag{4.17}$$

is still valid. The performance of this algorithm depends heavily on the choice of the initial seeds. Ideally, one seed per image region must be provided as it is described in section (4.5.4). Usually there is more than one seed per image region, and a merging procedure must be devised in order to merge adjacent regions that have similar properties. The homogeneity rule in this algorithm is based on fitting a planar surface to the original image data in the seed region and subsequent growth regions. The region growing process is controlled directly by the planar-fit error e_i, obtained at each iteration, and a pre-determined allowable tolerance as computed using equation (4.31). The iteration continues until the termination criteria are met at which point the computed planar region description is rejected or accepted. Each part of the region-based segmentation algorithm is described in detail in the following sections. The material is divided into main sections on image noise estimation, flat pixel type labeling, edge pixel extraction, seed region extraction and region growing. The algorithm works iteratively on every region of interest one at a time. It should be mentioned that in the presence of a high quality, high resolution DSM, such as one provided by a laser scanner (5 to 10 points per 1 m^2), the segmentation process can be performed on the DSM itself to directly extract the symbolic 3D planar-roof polygons.

4.5.1 Image Noise Estimation

In order to group pixels based on a planar surface fit in the region segmentation algorithm, it is necessary to know a priori how well the planar primitives should fit the data. This type of information should be derived from the image data in a data-driven mechanism, so that the algorithm can adapt to the noise conditions. This section describes a simple method for estimating both, the average noise in the entire image, called *global image noise*, and average noise within every region of interest called *local image noise*, with the assumption that the additive noise process is relatively stationary across the image. This method is a modified approach of Besl (1988), using the root mean squares error (RMSE), of a local planar surface fit. Better methods of measuring image noise are not doubt possible (Lemmens 1996, Waegli 1998), but good results were obtained with the following simple method. Global and local image noise estimates were found useful for indications of image noise variance and image quality.

Within every region of interest, perform a least squares planar fit $z = ax + by + c$, in a 3×3 neighborhood of every pixel, to compute the slope of each plane at that pixel. If the slope of the plane at a pixel is greater than a pre-specified slope-threshold, disregard the pixel since it is probably at or near to a step discontinuity. Similarly, if the slope of the plane is exactly zero, the neighborhood of the pixel is likely to be synthetic data, or completely dark, and it should be discarded. If the pixel has not been discarded, compute the planar RMSE fit for the pixel, σ_{pixel}, as defined by:

$$\sigma_{pixel}^2 = \frac{1}{9} \sum_{(x,y) \in win} (I(x,y) - (ax + by + cz))^2 \tag{4.18}$$

where $I(x,y)$, is the original digital image, and a, b, c, are the coefficients of the estimated plane. Compute the average mean of the fit error σ_{pixel}, for the pixel within the region of interest. This quality measure is called local image noise estimate $Local_e$ of, the corresponding region of interest and is defined by:

$$Local_e = \frac{1}{n_i} \sum_{1}^{n_i} \sigma_{pixel}(i) \tag{4.19}$$

where n_i is the number of pixels, which are not rejected within the region of interest i. The weighted arithmetic mean of local image noise $Local_e$ of all the regions of interest is called global image noise estimate $Global_e$, and is defined as:

$$Global_e = \sum_{1}^{m} n_i \cdot Local_e(i) / \sum_{1}^{m} n_i \tag{4.20}$$

where m, is the number of regions of interest. As it is described in the following sections, utilizing these two parameters enabled us to tie algorithm thresholds involved in our approach to the amount of noise in the image in an empirical manner.

4.5.2 Flat Pixels Type Labeling

A common approach to region segmentation is to start from some pixels (seeds) representing distinct image regions and to grow them, until they cover the entire image. The performance of this algorithm depends heavily on the choice of the initial seeds. Ideally, one seed per image region must be provided. The user in a supervised mode usually chooses the seeds. But, in order to implement a region growing segmentation algorithm that can be executed in an automated and unsupervised environment, a rule describing and extracting seed regions based on a data-driven mechanism is needed. To realize the concept, a strategy based on geometric characteristics of a digital surface has been developed. This approach is originally proposed by Besl and Jain (1988), where the sign of mean and Gaussian curvatures have been used to initially classify range images in industrial application into eight different surface types. These surfaces are graphically illustrated in figure (4.2) and are tabulated based on the sign of the surface curvatures H and K, in table (4.2). Moreover, an iterative region growing algorithm based on variable-order surface fitting has been utilized to partition the range image into smooth and meaningful surface regions. The essential difference of our approach compared to Besl is that based on the assumption that building roofs are composed of generic planar surfaces, we only concentrate on *flat surface type pixels*, which serve as the seed regions for the region growing segmentation process. In addition, the segmentation process is performed within the extracted regions of interest in aerial image based upon intensity grey values, not the height data in range image.

	$K > 0$	$K = 0$	$K < 0$
$H < 0$	peak	ridge	saddle-ridge
$H = 0$	(none)	flat	minimal
$H > 0$	pit	valley	saddle-valley

Table 4.2: Eight basic surface types defined by mean and Gaussian curvature signs (courtesy of Besl 1988)

To compute surface curvatures from digital images, the five partial derivatives $S_u, S_v, S_{uv}, S_{uu}, S_{vv}$, are all we need to compute the six fundamental form coefficients E, F, G, L, M, N (see equations 4.5, 4.8), and hence the mean and Gaussian curvatures by:

$$H = \frac{S_{uu} + S_{vv} + S_{uu}S_v^2 + S_{vv}S_u^2 - 2S_uS_vS_{uv}}{2(1 + S_u^2 + S_v^2)^{3/2}} \tag{4.21}$$

$$K = \frac{S_{uu}S_{vv} - S_{uv}^2}{(1 + S_u^2 + S_v^2)^2}. \tag{4.22}$$

The problem to be addressed here is computing these partial derivatives through the given digital image. In fact, they have to be replaced by their approximations computed from the discrete surface. A possible solution is based on a local least squares surface model using discrete orthogonal polynomials which has been discussed extensively in (Bolle & Cooper 1984, Haralick 1984, Besl 1986). The least squares estimates of first and second partial derivatives are determined based on the following separable binomial window operators:

$$\mathbf{D}_u = \vec{d}_0\vec{d}_1^T \quad , \quad \mathbf{D}_v = \vec{d}_1\vec{d}_0^T \tag{4.23}$$

$$\mathbf{D}_{uu} = \vec{d}_0\vec{d}_2^T \quad , \quad \mathbf{D}_{vv} = \vec{d}_2\vec{d}_0^T \quad , \quad \mathbf{D}_{uv} = \vec{d}_1\vec{d}_1^T$$

where the column vectors for a e.g., 5×5 window operator are defined as:

$$\begin{aligned} \vec{d}_0^T &= \frac{1}{5}[\ 1 \quad 1 \quad 1 \quad 1 \quad 1\ ] \\ \vec{d}_1^T &= \frac{1}{10}[\ -2 \quad -1 \quad 0 \quad 1 \quad 2\ ] \\ \vec{d}_2^T &= \frac{1}{14}[\ 2 \quad -1 \quad -2 \quad -1 \quad 2\ ]. \end{aligned} \tag{4.24}$$

Note, that estimated surface curvatures are extremely sensitive to noise because they require the approximation of second derivatives, in which high frequency noise is amplified. Thus, a smoothing operation is required. One way to filter out the local fluctuations in digital images $I(x, y)$, is to perform a morphological opening followed by a closing operation with a constant-valued structuring element k, which is somehow a *median filter* (equation 4.25). The net result of these two operations removes or attenuates both bright and dark artifacts or noise while leaving the edges sharp (Haralick & Shapiro 1992).

$$\bar{I} = (I \circ k) \bullet k \tag{4.25}$$

Therefore, the partial derivatives are computed utilizing the intermediate smoothed image and the least squares derivative window operators based on an image convolution as defined by:

$$I_u(u,v) = \mathbf{D}_u \star \bar{I}(u,v) \quad , \quad I_v(u,v) = \mathbf{D}_v \star \bar{I}(u,v) \tag{4.26}$$

$$I_{uu}(u,v) = \mathbf{D}_{uu} \star \bar{I}(u,v) \quad , \quad I_{vv}(u,v) = \mathbf{D}_{vv} \star \bar{I}(u,v) \quad , \quad I_{uv}(u,v) = \mathbf{D}_{uv} \star \bar{I}(u,v).$$

These estimated partial derivatives can then be plugged into the equations (4.21) and (4.22), to compute mean and Gaussian curvatures for each pixel of a digital image. Furthermore, the signs of mean $sign_H$, and Gaussian curvatures $sign_K$, are determined for each individual pixel based on two pre-specified tolerances $t_H = (min^H, max^H)$, $t_K = (min^K, max^K)$ and output of the thresholding functions (4.27, 4.28) respectively. In practice, this thresholding process is needed due to the presence of noise in the image and approximation errors caused by the computation of the required partial derivatives.

$$sign_H = 0 \quad if \quad min^H < H < max^H \tag{4.27}$$

$$sign_K = 0 \quad if \quad min^K < K < max^K \tag{4.28}$$

It is described earlier that we are only interested to planar surfaces, where values of both $sign_H$, and $sign_K$ are zero (table 4.2). In order to classify this type of regions in the image, within every region of interest, flat-pixels type are labeled and stored in an image called *flat-pixel type binary image*. This binary image is in fact, a coarse classification of the image and is used as an input to the seed region extraction algorithm, which is discussed in the following section.

4.5.3 Edge Pixels Extraction

Having the values of first partial derivatives of each pixel (I_u, I_v), an approximate measure of the edge magnitude can be computed in an inexpensive manner using (4.29).

$$edge_{mag} = \sqrt{I_u^2 + I_v^2} \tag{4.29}$$

A threshold on this edge magnitude provides a detector for edge pixels. In current implementations this feature has been used to compute the *edge magnitude image*. This image can then be thresholded based on the estimated noise variance in the image, to compute an *edge pixels binary image* (figure 4.6). In addition, the edge magnitude image is used as a weight function for the selected edge pixels during the hypothesis model verification process discussed in chapter 6.

The entire computation is based on a simple fact: if the value of a pixel in the edge magnitude image is more than a pre-specified threshold t_{edge}, then this pixel is an edge pixel. The t_{edge} is an edge threshold and is defined for each region of interest by :

$$t_{edge} = C \times Region_e \tag{4.30}$$

where C is a constant value, and $Region_e$ is the output of the following function:

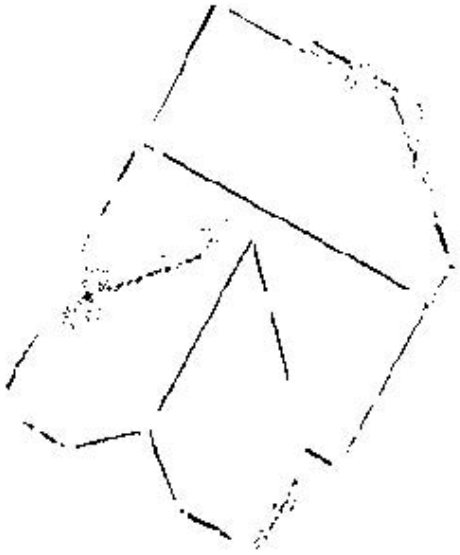


Figure 4.6: Computed edge pixels binary image.

$$Region_e = Min(Local_e, Global_e). \tag{4.31}$$

The above function will accept the local image noise and global image noise values (section 4.5.1) as input and assign the minimum value to the $Region_e$ for each region of interest.

The computed binary image is an input to the segmentation algorithm and serves as a constraint data structure during region growing. This is elaborated in section (4.5.5.2).

4.5.4 Seed Region Extraction

Flat pixels type of one primitive roof region tend to connect to another neighboring, but distinct primitive roof. To overcome this problem it is proposed that connected regions should be isolated and eroded until a small, maximally interior, single seed region is isolated. This seed region is then grown based on a planar surface fitting algorithm until it reaches its natural limits as defined by variations in the image data. Given the flat pixels type binary image, the following method has been utilized to extract a small interior and correctly labeled group of connected pixels, called *seed region*. The algorithm begins to isolate the largest connected flat pixels region using a 4-connected component analysis. The isolated region is eroded repetitively using a region erosion operator, after each erosion, there exist a largest 4-connected subregion of the original region. The largest subregion with the minimum number of pixels greater than or equal to a pre-specified minimum seed region size threshold (minimum number of seed region's pixels) is assigned to be a seed region. The minimum seed region size threshold must be equal or greater than the minimum number of points required for the planar surface fit. If the threshold is equal to the minimum number of points i.e., 3 pixels, then the planar fit can respond strongly to noise in the image. Therefore, the threshold should be greater than the minimum required number of pixels. Since the primary purpose of this strategy is to find a small enough isolated interior seed region that is not accidentally connected to any adjacent regions, and is far enough inside the boundaries of the actual surface primitive having escaped the undesired side effects of differentiation at surface boundaries, there is an upper limit on the number of necessary erosion. This limit is based on the window size of the derivative operators used in the computation of surface curvatures. In a 2D image convolution by a window operator, e.g. $L \times L$, a given pixel is affected by the input data $(L-1)/2$ pixel away from it on any side. Therefore, a limit of $(L+1)/2$ erosion iteration reduce a $L+1 \times L+1$ binary block to nothing implying that there are no effects of the $L \times L$ window operators after $(L+1)/2$ erosions.

The entire algorithm stops either the maximum number of erosions reached or when the minimum number of pixel in the largest four-connected subregion is greater than or equal to the minimum seed region size threshold.

Figure 4.7: Extracted seed region overlaid on corresponding region of interest

Figure (4.7) shows the extracted seed regions overlaid on a corresponding region of interest. This small region serves as the seed region to the iterative region growing algorithm.

4.5.5 Region Growing

As its name implies, *Region growing* is a procedure that groups pixels or subregions into larger regions. The approach starts with a seed region and from this seed the region is growing by appending to each seed those neighboring pixels that have similar properties, e.g. grey level, etc. For each border pixel of a seed region, we check if there is any pixel in its 3×3 neighborhood in the image that can be merged with the region to which the border belongs. The candidate pixel P must not be an edge pixel and its value must be close to the computed value of its correspondence in the grown region $\hat{I}(x, y)$, that means:

$$|I(x, y) - \hat{I}(x, y)| \leq threshold. \tag{4.32}$$

To illustrate the basic concepts of our approach let us start with the extracted seed region from the previous section. First, it must be decided how well a planar surface should fit the original data. The image noise estimation procedure discussed previously provides an indication of the maximum fit error threshold for the iterative surface fitting algorithm. A plane is fitted to the small seed region based on a least squares process. If the seed region belongs to parts of a roof that is not too highly curved, this plane will fit quite well to the original data. If the plane fits the seed region within the maximum allowable fit error threshold, then the seed is allowed to grow, if not, the seed is rejected.

After a plane is fitted to a region, the plane description is used to grow the region into a larger region where all pixels in the largest region are connected to the original region and are compatible in some sense with the approximating planar surface function for the original region. On the *n-th* iteration for the seed region S_i^0 corresponding to the primitive region R_i, the region growing algorithm accepts the smoothed original image $\bar{I}(x, y)$, the plane description $P\{(a, b, c); (x, y) \in R_i^n\}$ computed from least squares planar fitting algorithm, the edge pixels binary image and the surface fit error e_i^n computed from the fit to the primitive region R_i^n. The first step is to compute the absolute value of the vector ΔZ_i^n, given by:

$$\Delta Z_i^n = |\bar{I}(x, y) - \hat{I}(x, y)|. \tag{4.33}$$

The function $\hat{I}(x, y)$, is computed based on the planar function $P\{(a, b, c); (x, y) \in R_i^n\}$, for all the pixels of region R_i^n, and its neighboring pixels which are called in this context candidate pixels. Vector ΔZ_i^n has small values

if the original image surface lies close to the approximating plane and have large values otherwise. Therefore, this vector is thresholded to find the compatible pixel, which are defined as:

$$P_{compatible} = \{(x,y)|\Delta Z_i^n(x,y) \leq t_{pixel}\} \tag{4.34}$$

where t_{pixel}, is the error tolerance threshold that determines how close the candidate pixel must be to the approximating plane, so that they are considered compatible. The next region R_i^{n+1}, is computed based on these compatible pixels.

After the region growing iterations have terminated, one is left with the grown region R_i, along with the approximating plane parameters, and the fit error e_i. When a grown region is rejected, the seed region responsible for the grown surface region is marked off in a writable copy of the flat pixels type binary image as having been processed, which prohibits the use of the same seed region again. When a grown region is accepted, all pixels in the accepted region are similarly marked off in the flat pixels type binary image so that they are not considered for subsequent seed regions. In this respect, surface rejection and surface acceptance are similar. However, the surface acceptance process is much more complex in that it updates several other data structures such as the *error image* $e(x,y)$, the *best fit region label image* $Region_{label}(x,y)$, and the list of primitive plane-roof regions. Purpose and description of these data structures will be discussed next.

4.5.5.1 Error Tolerance Thresholds

There are two error tolerance thresholds used for the region-based segmentation algorithm t_{pixel} and E_{max}. The maximum allowable error fit threshold called *region threshold function*, E_{max}, is used to allow a region to grow or stops. Assuming a valid estimate of the noise variance in the image, i.e. $Region_e$, as determined in (4.31), the root mean square error of a planar fit e_i should also be the same value. But since this quality measure of the planar fit, itself is a random variable, there are variations in this quality measure from one planar fit to another. Therefore, the maximum allowable fit error threshold must be greater than the estimated noise variance in the image. So, if RMSE fit of a particular region on *n-th* iteration e_i^n will be less than the region threshold, then the error fit test is passed and the region is allowed to continue growing. Otherwise, the growing process stops and a grown region R_i^n is accepted as primitive roof region if its size is bigger than the pre-specified minimum primitive region size, if not, it will be rejected. There is a small deviation in this strategy for the first iteration, in this particular case if $e_i^1 \geq E_{max}$, then the corresponding seed region $S_i^0 = R_i^1$ is rejected outright, based on the assumption that the extracted seed does not belong to part of a flat roof. The region threshold function is defined by:

$$E_{max} = C_1 \cdot Region_e \tag{4.35}$$

where C_1 is a constant value, e.g. $C_1 = 2.8$, to increase the estimated noise variance in the image. The second quality measure, the error tolerance function t_{pixel}, is also a function that increases the value of the root mean square error fit e_i in each iteration n during the process of thresholding candidate pixels and is defined by:

$$t_{pixel} = C_2 \cdot e_i^n \tag{4.36}$$

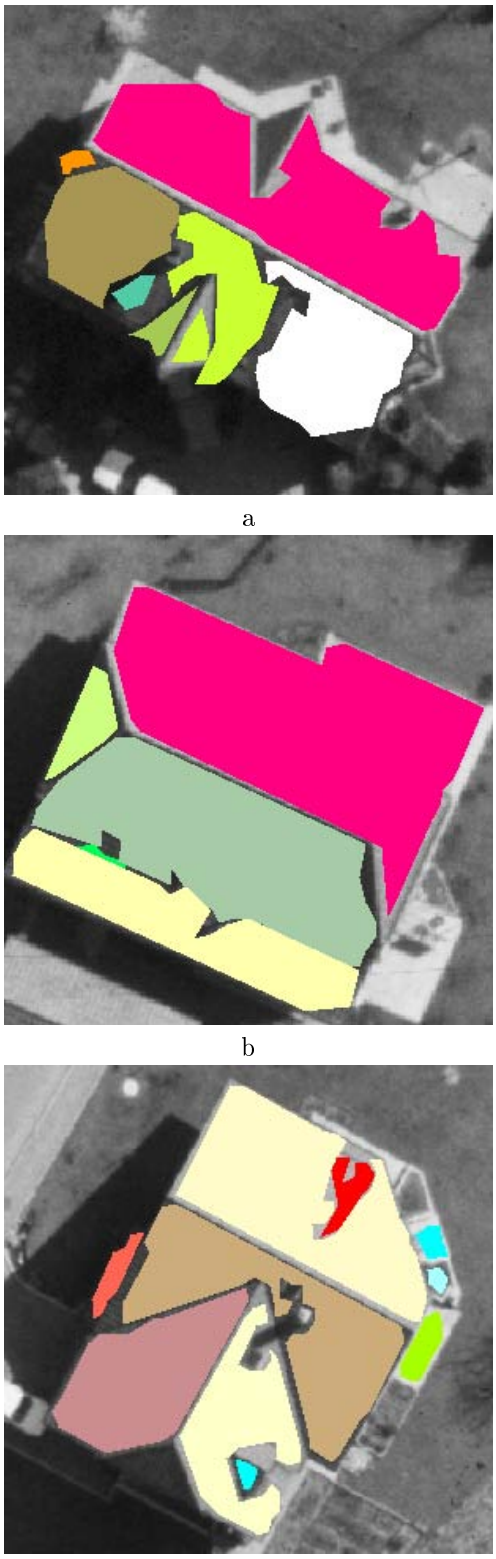
where C_2 is a constant value. Based on the assumption that the present noise is relatively stationary across the image, a value of e.g. $C_2 = 2.8$ offers a reasonable form for error tolerance function, because approximately about 99.5% of all the candidate pixels of the approximated planar region lie within this error tolerance. In this simple error tolerance function, the factor C_2 controls the aggressiveness of each region growing iteration and therefore, controls the speed and accuracy of the iterative planar fitting process.

4.5.5.2 Relaxation Labeling in Region Analysis

Most region segmentation methods are deterministic in the sense that they assign each image pixel to just one region. Although such a segmentation is ultimately desirable, it is not always useful to employ such segmentation during the growing process, because they treat ambiguous cases –pixels lying in transition regions–, in a rather inflexible way (Gonzalez & Woods 1993). The fundamental problem with applying this strategy is that outlier pixels picked up in one iteration can never be dropped in subsequent iterations. This implies that there is something wrong with the simple pixel compatibility requirements for new pixel. We apply a relaxation

methodology in our approach by introducing two new data structures, the error image $e(x, y)$ and best fit region label image $Region_{label}(x, y)$. These two images produce a mechanism that if oulier pixels enter a region on one iteration, they leave on a later iteration, depending on the results of the planar fitting. Hence, the effect of few bad pixel is not cumulative in this case. When a region iteration terminates and the primitive plane-roof region is accepted by successfully passing the test conditions discussed previously, the magnitude of the residual error at each pixel of the grown region is stored in the error image to explicitly note the spatial distribution of the approximation errors. For the region label i, each pixel of the accepted plane-roof region is stored in the best fit region label image $Region_{label}$, to explicitly note the pixel that the approximating plane fits to within the specified threshold. During the thresholding operation that forms the compatible pixel, each pixel must not only have an error less than the error tolerance threshold t_{pixel} (4.34), but it must also have an error less than the current error stored in the error image. If both conditions are not satisfied, then the pixel is not considered compatible with the growing region despite the fact that the allowable error tolerance requirement is met. The error image approach provides a relaxation capability for a pixel already associated with a given roof primitive as opposed to strictly forbidding the reassignment of a pixel to another roof region once they have been assigned to one roof primitive. That is, a previous plane-roof region may approximate the value at a pixel well enough for that pixel to be associated with it, but if the current region approximates the pixel value better, then the pixel can be relabeled with the current region's label if the pixel meets connectivity requirements with other pixels. In this manner, the error image behaves as a region growing constraint. The other region growing constraint, which has been used during the construction of the candidate pixel is that a region should not grow over step discontinuities or orientation discontinuities. This constraint is simply applied by integrating the extracted edge information with the region growing algorithm. A pixel is considered a candidate pixel in a 3×3, neighborhood of a border pixel, if it is an off-pixel in the edge pixels binary image.

Figure 4.8 illustrates the results of extracting 2D plane-roof regions for three different roof structures. It shows that main structures of the roofs are correctly extracted. However because of the presence of noise in the image and due to the shadow caused by the microstructure on top of the roof (4.8-a), some of the larger roof primitives are divided into the smaller primitives. In fact, these intermediate 2D regions are merged into the larger one, if they satisfy the compatibility requirements during the reconstruction process, which is the subject of the next chapter. In addition –in the presence of high quality, high resolution digital images–, the proposed method is capable to detect and extract the microstructure on top of the buildings roof such as dormer windows in figures (4.8-a) and (4.8-c). This type of information improves significantly the results of the higher-level reconstruction processes, in particular when dealing with complex buildings. Moreover regional information derived from the extracted regions provide many descriptions such as area, surface normal, etc., which are not derivable from an edge- or line-based segmentation algorithm. They also provide topological information based on the computation of a *Polygons Adjacency Relationships (PAR)* (see section 5.3), which is an essential requirement for combining the simple, image-oriented geometric primitves in a lower-level process to more complex, model-oriented geometric primitives or structures in a higher-level process.



a

b

c

Figure 4.8: Extracted 2D plane-roof regions overlaid on corresponding buildings roof: a) gable roof structure building, b) hipped-gable roof structure building, .c) complex roof structure building.

Chapter 5

Generic Polyhedral-Like Model Reconstruction

5.1 Introduction

An automated vision system must be able to determine the appropriate transition from the more image-oriented, qualitative representation of the object in the lower levels to the more abstract model-oriented, quantitative representation of the object at the higher levels. A major problem in precise definition of the nature of the mapping is the modeling aspect. Creation of definitive models is difficult due to enormous variations in the geometric and functional descriptions of the objects of interest. In addition, the embedding of an object in a scene and the imaging process itself may introduce many different kinds of noise and distributions. Objects may be partially occluded by other objects, the scene may have particularly high contrast, the sensor may be particularly noisy, and so forth.

This chapter describes a new method for automatic 3D reconstruction of polyhedral-like objects, in this context used as a generic building model. A boundary representation of a coarse building hypothesis is constructed in a data-driven, bottom-up approach, from simple geometric primitives (2D plane-roof regions) in image domain to more complex geometric model (3D-roof structure) in object domain.

The previous chapter introduced the main aspects of the recognition task. The purpose and strategy used in subsequent low-level processes to detect and extract the 2D plane-roof regions which have meaningful correspondence with the roof components of the building objects are discussed. Consequently, it is now time to move to the more model oriented representation of the buildings, which is carried out in the reconstruction part of this study and is the subject of this chapter. The reconstruction procedure consists of different intermediate, interrelated processes aiming to form a framework, in such a way that every process provides more abstract and more object related information to its immediate higher level process. This chapter is organized in two parts, the first part describes several mid-level vision processes, starting with estimating the preliminary parameters of the 3D polygonal primitives of roof structures, called *3D plane-roof polygons* in this study, by back projecting the corresponding extracted 2D plane-roof regions in image space into the 3D object space. This process is performed based on a synthesis robust parameter estimator developed in chapter 3 and is discussed in section (5.2). To topologically describe interrelations between these 3D geometric primitives, which is an essential requirement for an automated reconstruction process, the *Polygons Adjacency Relationship (PAR)* is computed (section 5.3). These adjacency relationships are defined based on Voronoi diagram (dual of Delaunay triangulation) and describe the topological properties, in particular the neighborhood relationships between the basic elements of a roof structure. Based on the computed PAR the compatible adjacent 3D polygons are merged into the larger 3D plane-roof polygon. Its symmetry with respect to their adjacent polygons is also defined and stored as attributes for further processes. This merging process is covered in section (5.4). The primitive 3D elements along their adjacency relationships information and derived attributes are input to the POLY-MODELER, where they are geometrically and/or topologically combined to generate the coarse building model. The POLY-MODELER is a new generic polyhedral-like model generator, which is originally developed in this study (Ameri & Fritsch 1999). It is based on a generic polyhedral-like solid model and generates the boundary representation (b-rep) of a coarse hypothesis building model using the 3D intersection of adjacent polygons. The second part of the current chapter (section 5.5) is dedicated to the mathematical concept, notations and a detailed discussion of POLY-MODELER. The proposed methods and the mid-level vision processes discussed in this chapter are all implemented and the subsequent results of different processes are presented.

5.2 Primary Roof Elements in 3D Object Space

An automated vision process, such as 3D object reconstruction, can be described as a complex mapping function to transfer the mass of low-level image-domain measurable knowledge into the more abstract form of high-level object-domain semantic knowledge of the world's object. An important issue in this complex procedure is an early selection of relevant knowledge in lower domain, and entering the higher level process as early as

possible, so that the deduction processes do not become combinatorial, thus reducing the search complexity and computational expenses (see section 4.2). In fact, in this transition, an intermediate level is required, where the initial extracted knowledge in the image space, which is represented in the more compact symbolic description of image-driven primitives, e.g. 2D regions, can be integrated or transferred into the more symbolic representation of the object-driven primitives, e.g. 3D planar polygons. As a result, initial hypotheses of the essential components and elements of the object of interest are created. Therefore, immediately after detection of the primary 2D plane-roof regions they should translate into the corresponding 3D plane-roof polygons, simply by inversion of the imaging process. Having an initial description or approximation of the object surface in real world, i.e. in this study the corresponding DSM, the transformation process is equivalent to a 3D regression problem and is performed using standard collinearity equations (6.10). Theoretically, the best result is obtained based on the traditional least squares fitting solution. However, in practice, as it has already been pointed out in chapter 3, due to the presence of outliers in the original data, the solution is far more complex than the simple fitting process and an appropriate robust fitting procedure is required. Outliers are pervasive phenomena in vision theory and are emerging differently in every vision application depending on the practical situation (Schunck 1990). Particularly, in this application outliers occured in both, DSM and extracted primary 2D regions. Outliers appearing in the extracted 2D regions are caused by the failure of the segmentation procedure. As it is shown in figure (3.4-a), due to the presence of noise, shadow or low contrast scene in the image, as well as the nature of a region-growing based segmentation algorithm, the segmentation process sometimes grows over the discontinuities or the physical bounding edge of the region. This segmented region part(s) is appearing as outliers during the fitting process, as its real height is significantly lower than the height of the corresponding building roof, and therefore leading to an arbitrary solution, if outliers are not detected and excluded from the estimation process, as it is illustrated in figure (3.4-b). Outliers occurring in the DSM are caused during its generation. In fact, commonly there are two different comparable methods for the generation of DSM. 1) Automatic photogrammetric techniques such as least squares (area-based) matching (Ackermann 1984, Förstner 1982), or feature-based matching algorithms (Förstner 1986). 2) Airborne laser scanning methods such as continuos wave, or pulse techniques (Wehr & Lohr 1999). Both methods have advantages and disadvantages, a complete comparison of various aspects of these techniques is given by (Baltsavias 1999). Despite the fact that the laser scanning technique provides high accurate direct geometric descriptions of visible surfaces, practical result in the particular application of automated building reconstruction has shown that the quality of the DSM in built-up areas generated by either technique is still insufficient. For example, issues like occlusion, shadow and anomalies of the surface height and discontinuities in photogrammetric methods, and lack of explicit measurement of breaklines such as roof ridges in laser scanning techniques, as well as subsequent smoothing operations, thus trim off the roof corners or the roof parts are caused that both methods failed to accurately recover the descriptions of the roof structures. In fact, the latter method is capable of partially overcoming this problem by highly dense sampling measurements of the terrain surface, which in this case ask for very expensive and costly operations. The effects of the presented outliers are eliminated in two steps in this research. The extreme ouliers, which lead to an arbitrary plane parameters for the 3D plane-roof polygons are detected and excluded during the regression procedure using a synthesis robust parameter estimation developed in chapter 3 and is discussed shortly in this section. The reminder of the outliers that cause a minor deviation between the estimated roof structure parameters and the physical ones are eliminated during the verification of hypothesis roof structure, which is discussed in the next chapter.

Recall from section (3.4), the proposed robust parameters estimation method is a two-stage parameters estimation algorithm. The first stage flushes and detects the outliers and estimates the best initial 3D plane parameters based on the inliers data using a random sampling type estimator such as Random Sampling Consensus *RANSAC* (Fischler & Bolles 1981), or alternatively Least Median Squares, *LMS* (Rousseeuw & Leroy 1987). The estimated parameter values along with the estimated error variance are then introduced into the iterative re-weighting M-estimator algorithm (Huber 1981, Hampel et al. 1986), as initial values to compute the final 3D plane parameters.

There are two important issues that have to be discussed. Firstly, before transferring the extracted 2D plane-roof regions, which are stored in a raster-based data structure, into the 3D object space, they are converted into the 2D plane-roof polygons and stored in a vector-based data structure. That is why the corresponding 3D primitives in object space are called *3D plane-roof polygons*. However, both representations along a direct one-to-one correspondence are stored in the system as they are required in subsequent processes. For example, the raster-based structure is used during the computation of the polygon adjacency relationships (PAR). Secondly, during the fitting process, despite the fact that bounding vertices of the 2D polygons are sufficient to estimate the 3D plane parameters. A pre-specified number of the pixels e.g., 200 pixels, within the extracted 2D regions are sampled in a regular interval and introduced into the regression process as observations, in order to increase the redundancy of the observations and therefore the reliability of the solutions.

Figure (5.1) illustrates the results of transferring the extracted 2D plane-roof regions into the corresponding 3D plane-roof-polygons in object space for a gable (5.1-b), a hipped-gable (5.1-d), and a complex (5.1-f) roof structure of the residential buildings of Avenches data set. It illustrates that the algorithm correctly recovers the parameters of the 3D plane, even in the presence of the disturbances such as shadow or small microstructure on top of the buildings roof (figures 5.1-b and 5.1-f).

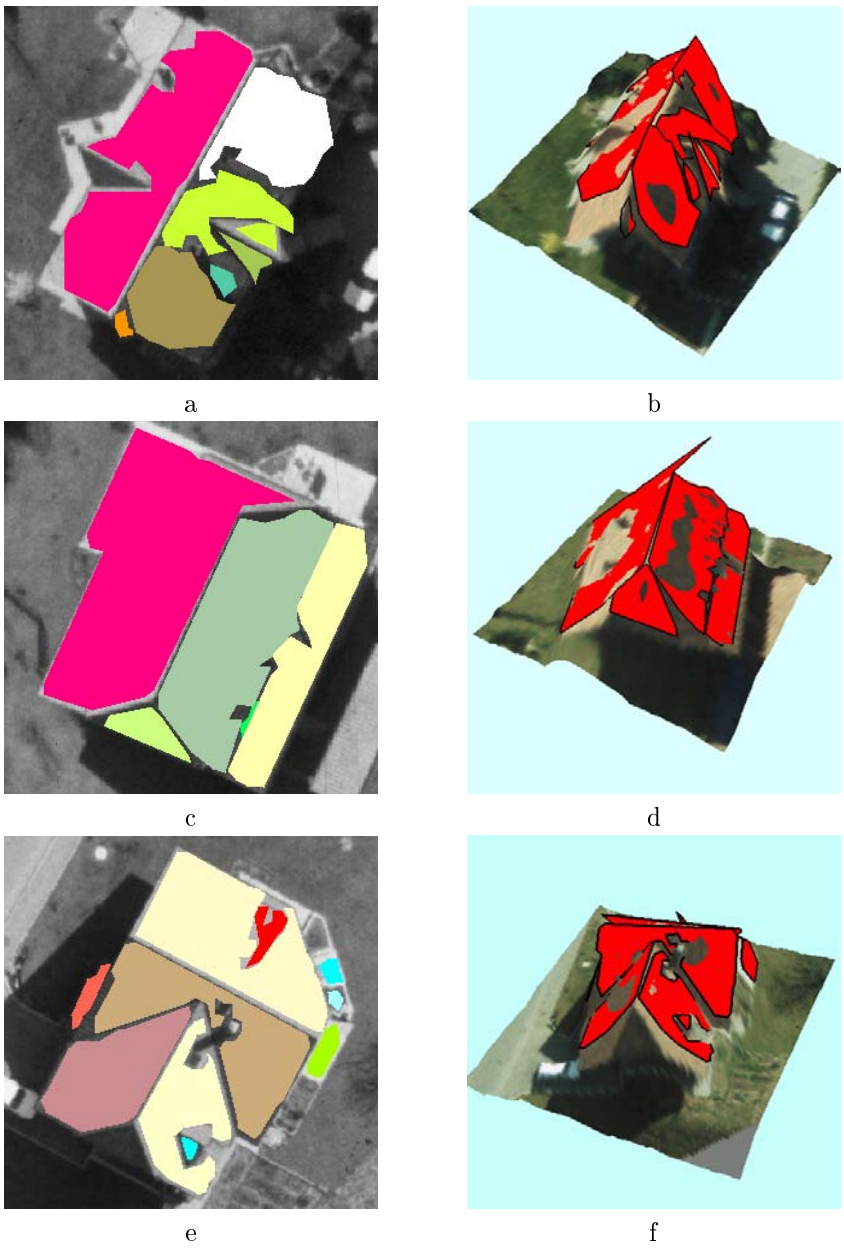


Figure 5.1: Estimated 3D plane-roof polygons in object space: a) extracted 2D plane roof regions of a gable roof structure, b) estimated 3D plane-roof polygons overlaid on corresponding gable roof structure building in object space, c) extracted 2D plane roof regions, and d) estimated 3D plane-roof polygons overlaid on corresponding hipped-gable roof structure, e) extracted 2D plane roof regions, and f) estimated 3D plane-roof polygons overlaid on corresponding complex roof structure building.

5.3 Polygons Adjacency Relationships

The basic idea in geometric modeling is to combine simple shapes to construct complex models. In our particular modeling process of 3D building reconstruction, the adjacent 3D plane-roof polygons, shortly *3D-poly*, are combined to construct the roof structure. Modeling complex objects such as buildings requires considerable attention to their topology. We must understand how simple elements are connected to form the complex model and how its topology is preserved when subjected to a variety of transformations. Topological properties are not metrical, but concern such things as connectivity and dimensional continuity (Mortenson 1997) [1]. In the previous section we discussed processes of construction of the primitive 3D plane-roof polygons, which are the main components of the roof structure and form the foundation on which we build the generic building models. In order to topologically describe the interrelation between these 3D primitives, a *Polygon Adjacency Relationship (PAR)* is computed. The concept of spatial adjacency, which has been normally defined based on a point-wise data set, is extended by introducing the adjacency relationships between polygonal primitives of different shapes and sizes, including connected, disconnected, or overlapped ones. During the reconstruction process, the PAR provides the essential topological information such as adjacency, and 'contained-in' relationships between incorporated primitives, which are the minimum types of object relationships that are required in an automated vision process based on a generic object model. The PAR is defined based on a Voronoi diagram (dual of Delaunay triangulation), where each primitive plane-roof polygons, in this context a data point, produce a zone of influence representing all parts of the space closer to that polygon than to any other. Polygons are considered adjacent only if their Voronoi regions touched. In fact, the main reason for using Voronoi regions for solving the problem is that no model of spatial adjacency is available for disconnected objects, and hence the definition of adjacency had to await for the connection of the points, line segments or polygons in the form of a graph structure by techniques that are primarily coordinate-based line intersection detection methods (Gold 1990). The Voronoi diagram and the Delaunay triangulation are closely related, and one can be extracted from the other. There are several algorithms for the generation of Delaunay networks and Voronoi diagrams (Midtbø 1993). In this section a description on the extended method of PAR computation developed in this research study is presented, and a discussion which reveal the importance of the concepts, and the methods of their computations that we have touched on only briefly here.

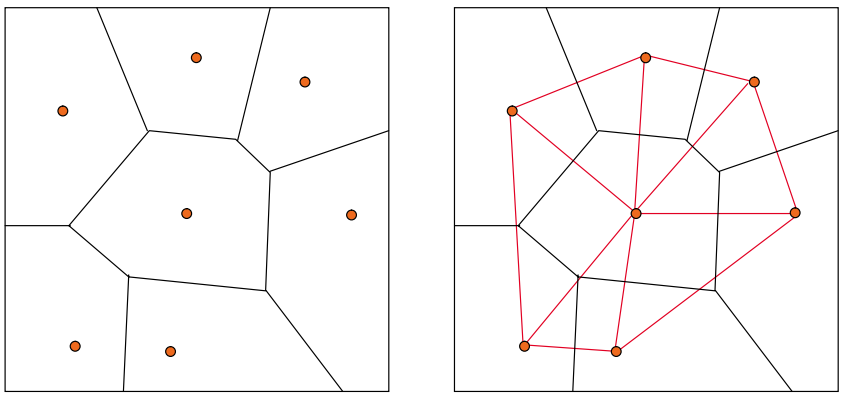


Figure 5.2: Correspondence between a Voronoi diagram and its dual, Delaunay triangulation network, a) Voronoi diagram, b) corresponding Delaunay triangulation network

Figure (5.2-a) shows the Voronoi diagram computed for a set of randomly distributed points. The lines that bisect the lines between a center point and its surrounding points define a single Voronoi polygon. The bisecting lines and the connection lines are perpendicular to each other. When we use this rule for every point in the area, the area will be completely covered by adjacent polygons. Notice that the polygons on the boundary of the area are open, because they have no neighboring points in that direction. The dual of the Voronoi diagram is Delaunay triangulation. If the Voronoi diagram is used as a basis, the Delaunay triangulation can be constructed by drawing the lines between the points in adjacent polygons. When the construction is finished we have got a triangular network that covers the whole area. The relationship between the Voronoi diagram and corresponding Delaunay network is shown in figure (5.2-b). The Delaunay triangulation network can also be computed directly,

[1] The properties of geometric shapes that are invariant under transformations that strech, bend, twist, or compress a figure, without tearing, puncturing, or inducing self-intersection, are topological properties

based on the Delaunay criterion of empty circle property, *a Delaunay triangulation network consists of non-overlapping triangles where no points in the network are enclosed by the circumscribing circles of any triangle.* The circle centers are recognized as the vertices of the Voronoi diagram. This observation can be used to make algorithms for the generation of Voronoi diagrams based on Delaunay triangulation.

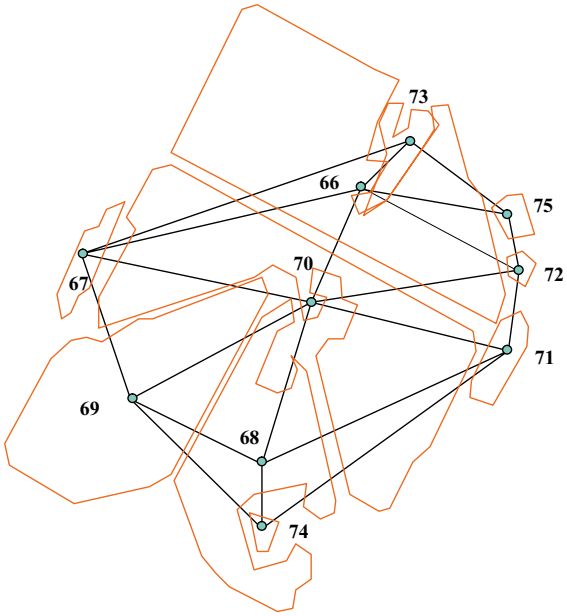


Figure 5.3: Delaunay triangulation network generated based on the central gravity points of corresponding 2D plane-roof polygons

An important property of Delaunay triangulation network is the adjacency relationship. If a Delaunay arc connects two points, then their associated Voronoi regions are adjacent to each other, and vice versa. The point adjacency relationship is the basis on which adjacency relationships concerning other geometric data types and features such as lines or polygons are defined. Principally, a Voronoi region can be constructed around any objects or geometric primitives in 2D space, which in turn gives us the ability to construct the adjacency relationships between more complex object types than the simple point-wise data set. In the application of computing the adjacency relationships between plane-roof polygons, due to the presence of polygons with different sizes, closeness, and polygons which are partially or totally overlapping each other such as 'contained-in' polygons, applying Delaunay triangulation method in vector domain causes undesired results. Figure (5.3) shows the generated Delaunay triangulation network between the set of points which are the center of gravity of their associated 2D plane-roof polygons. A close look at the figure reveal that e.g., polygon 67 is incorrectly adjacent to polygon 73, or polygon 74 is adjacent to polygons 69, and 71, which are not the valid adjacent relationships. Although, theoretically this type of problem can be solved applying the *constrained Delaunay triangulation criterion* [2] (Preparata & Shamos 1985, Midtbø 1993), where all the vertices of each polygon are considered as the data points instead of the corresponding center of gravity point, and the bounding edges of every polygon are introduced in the computation as pre-specified triangle edge constraints. However, in practice this type of constraint is also invalidated due to the presence of overlapping polygons, which are generated based on the nature of relaxation strategy in our region-growing algorithm, or the existence of the contained-in polygons i.e., polygons correspond to dormer windows on top of roof structure

In order to overcome this problem, the Voronoi diagram is generated based on *distance transformation* (Borgefors 1986), in a raster domain (Tang 1992). Pilouk et al. (1994), and Chen et al. (1994) have extended the concept into the 3D space for generating Delaunay tetrahedral tessellation. A distance transformation converts a binary image consisting of feature (kernel point), and non-feature pixel, into a gray-value image. The distance transformation operation assigns a number to every non-feature pixel, which is the distance between the corresponding pixel and the nearest kernel point. Computing these distances in digital images is based on an approximation of true Euclidean distance.

[2] A constrained Delaunay triangulation network is an extension of the standard method by allowing the pre-specified, non-intersecting line segments –except at their end points– to be forced in the computation as part of the triangulation network. The triangles containing any of such pre-specified edges may not be Delaunay triangles

There are several methods and associated masks for approximating the true Euclidean distance. Borgefors (1986) has discussed the performance results of applying different approaches in term of speed and maximal error between the true Euclidean distance and its approximation. She has pointed out that the problem of choosing the best distance transformation is application oriented. For an application that highly accurate result is not required, such as generation of Voronoi diagram for the purpose of obtaining the adjacency relationships between inexact features, i.e. extracted 2D plane-roof polygons from a noisy aerial image, computing exact distances from inexact features is not necessary, at least not when the exact distances are more computationally costly than adequate approximations.

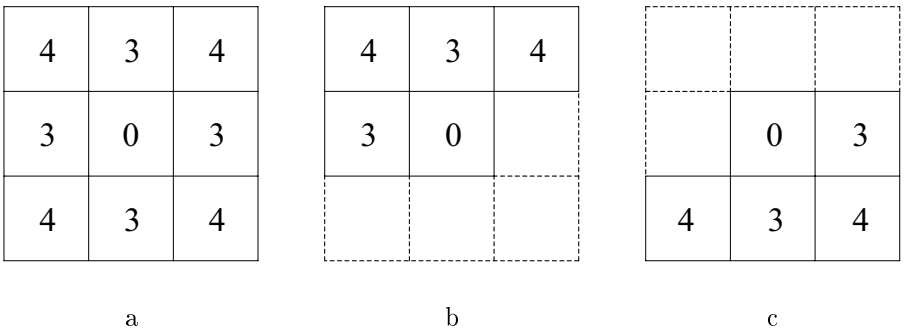


Figure 5.4: Chamfer 3-4 mask proposed by Borgefors (1986), a) symmetric Chamfer 3-4 mask used for parallel process, b) Chamfer 3-4 forward and , c) Chamfer 3-4 backward masks used for sequential process

Borgefors (1986) suggested the *Chamfer 3-4 mask* (see figure 5.4-a) for generation of Voronoi regions. Computing the distance values from all the data points using Chamfer mask, while at the same time keeping track of from which kernel point the distance is computed. This is done by initializing two new images the same size as the original image, the Voronoi, and the distance images. The new Voronoi image is used for tracking and is initialized by assigning the zero values to all its pixels except the corresponding kernel point pixels, which are marked by a unique number equivalent to the kernel point label. The distance image is initialized by assigning the highest integer value to all its pixels except the corresponding kernel point pixels, which are assigned to zero. The process may perform in a sequential procedure (Pilouk, Tempfli & Molenaar 1994). The symmetric Chamfer 3-4 mask is split into two *forward*, and *backward* masks as illustrated in figures (5.4-b), and (5.4c), respectively. The masks are passed over the image once each. The forward mask starts from upper-left corner of the image to the lower-right, and the backward mask scans the image from lower-right to the upper-left corner. At each pixel position, the minimum value of the sum of the distance image pixel values and the corresponding local distances of the mask is selected and assigned as a new value for the pixel in the distance image. At the same time the pixel in the Voronoi image that corresponds to the pixel which gets the new value in the distance image is marked with the label of the kernel point of which the distance is computed. The process is continued until all the pixels are scanned. After these two passes are performed, the distance image represents the distance transformation image of the kernel points, and all the pixels in Voronoi image that have the same label represent the Voronoi region corresponds to the kernel point. Figures (5.5-a), (5.5-b), and (5.5-c) shows the initial kernel points, the corresponding distance image, and the generated Voronoi diagram. In fact, two kernel points or features are adjacent if the associated regions are touched. As it was discussed previously, connecting the adjacent kernel points generates the Delaunay triangulation network.

The analysis of the result obtained by applying different geometric primitives and features as kernel points draw the important fact that the shape and size of the utilized features influence the result of the adjacency relationships significantly.

The figures (5.5-a) and (5.5-d) represent the same geometric features shown in figure (5.5-g) with different shapes and sizes. The Voronoi diagram computed based on the these data sets as initial kernel points are illustrated in figures (5.5-c), (5.5-f), and (5.5-l) respectively. As it is discussed above the result shows different adjacency relationships. In order to overcome this problem, the proposed method has extended in such a way that shape and boundary of the polygons are also taken into account. In fact, the principal of computation is the same only the initialization of Voronoi and distance images is different. Following is a stepwise summary of the extended method, which is applied for each region of interest or candidate building once a time, therefore reducing the required computational time.

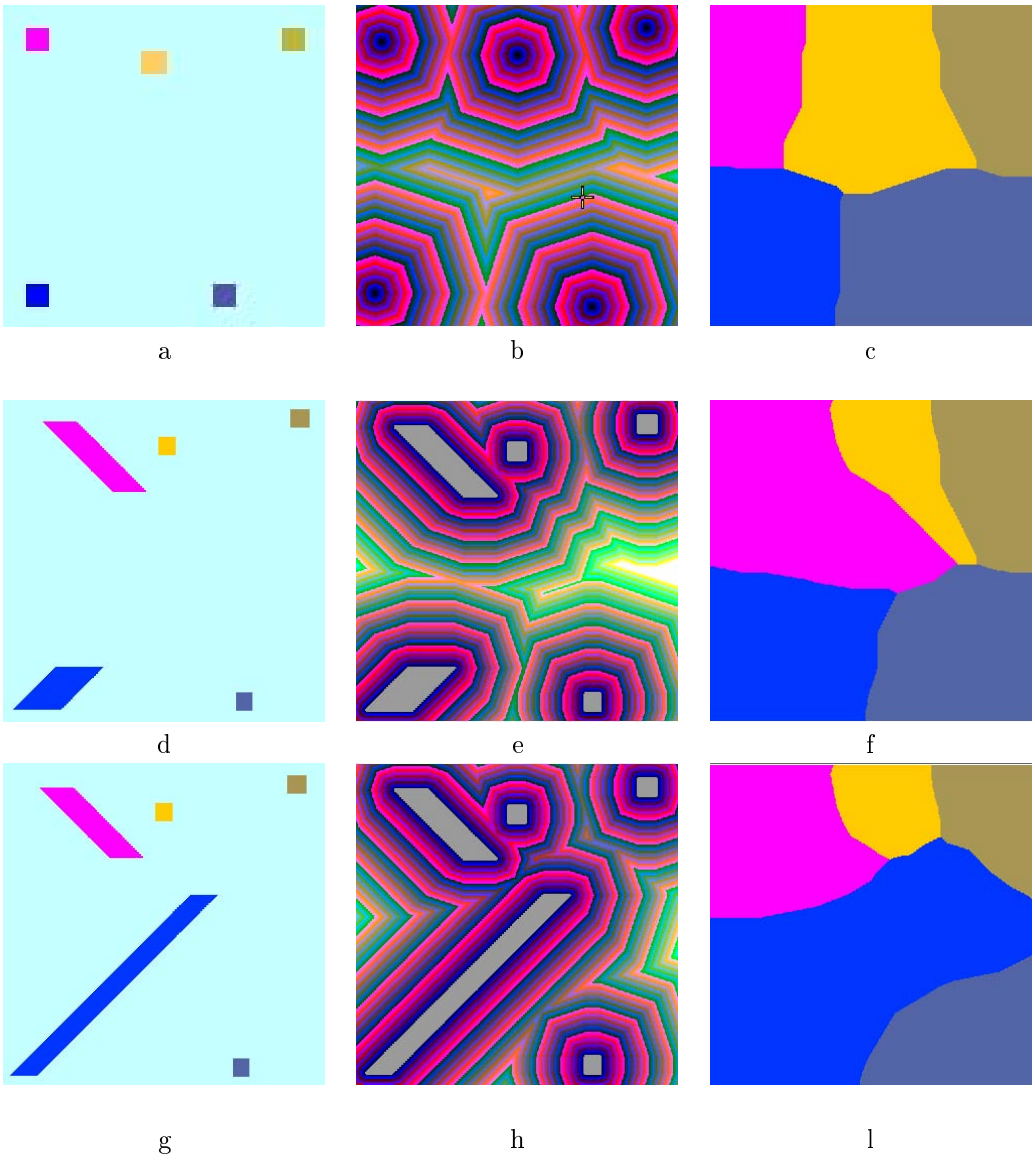


Figure 5.5: Computation of Voronoi diagram based on distance transformation, a) the central point of each polygon primitive is considered as kernel point, b), and c) show the corresponding distance image and Voronoi diagram respectively, d) the polygonal primitives (medium size) are considered as initial kernel points, e), and f) illustrate the corresponding distance image and Voronoi diagram, the adjacency relationships between primitives are changed, g) the complete polygonal primitives are considered as initial kernel points, g), and h) illustrate the corresponding distance image and Voronoi diagram, the adjacency relationships between primitives are significantly changed.

- Create two new images as Voronoi and distance images large enough to contain the corresponding region of interest.
- Make an ordered list of 2D plane-roof polygons based on their size, starting with the largest polygon.
- Initializes the images as discussed above. Note that instead of only initializing the central point of each polygon and inserting the bounding edges as external constraints (Tang 1992, Pilouk et al. 1994), all the corresponding pixels of the 2D plane-roof polygons are initialized in the images. The larger polygons are initialized first. In this manner, the smaller or 'contained-in' polygon pixels are not relabeled with the larger one.
- Perform the distance transformation and generate the Voronoi diagram as discussed above.
- Compute the polygon adjacency relationships (PAR), and store the result for subsequent processes.

The result of the proposed method for PAR of the extracted 2D plane-roof polygons within a region of interest is illustrated in figure (5.6). The generated Voronoi diagram (5.6-c) represents the adjacency between polygon primitives, in particular, adjacency relationships of the 'contained-in' polygons are correctly defined. The PAR is stored as complementary properties of each polygon primitives and updated in proceeding processes as required.

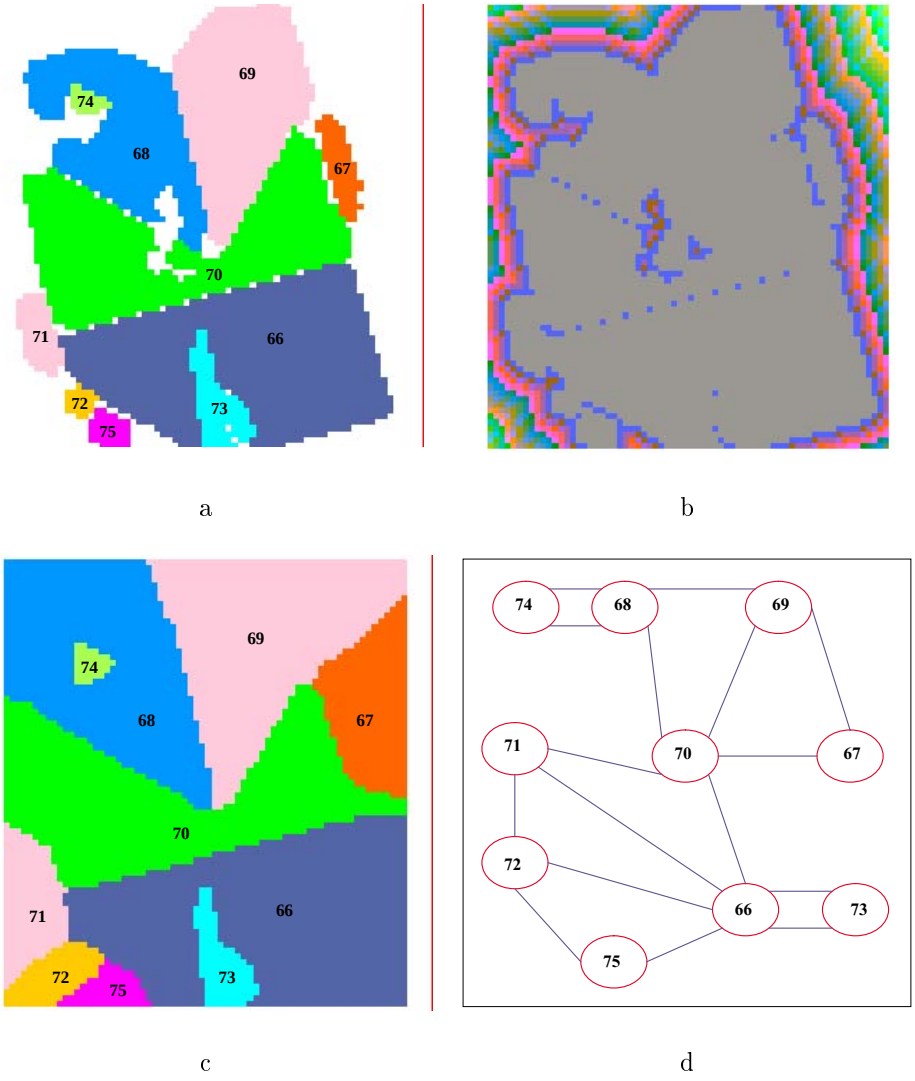


Figure 5.6: Computation of polygon adjacency relationships (PAR), a) initial kernel polygons, b) distance image, the gray area in the middle indicates the zero distance, c) generated Voronoi diagram, d) computed adjacency graph

5.4 Merging Compatible Adjacent 3D Polygons

It was discussed in the previous chapter that due to the presence of the noise, shadow or occlusion caused by the microstructures on top of the building roof some of the larger roof primitives are divided into the smaller primitives. Therefore, a merging procedure must be devised in order to merge adjacent intermediate polygons into the larger polygon primitives, if they satisfy the compatibility requirements. In fact the merging process can be done in image domain, during or immediately after segmentation process, where the similarity criteria are based on local statistical properties derived mostly from image intensity. It has been said here, and elsewhere (Fua & Leclerc 1990), that in a data-driven reconstruction process, object structures cannot be detected solely on the basis of the photometry information and methods based on purely local statistical criteria are bound to errors, thus geometric information obtained from the problem description, i.e., 3D planar surfaces, should be incorporated into the solution. In other words, if we have had enough relevant knowledge during segmentation process in image domain, we would not have the fragmented polygon primitives at that level. This was a sign to

approach this problem at the higher-level object domain where relevant object-related knowledge is available, that is 3D planar polygon primitives. The compatibility rules are thus, defined based on the descriptions of the planar polygons in 3D space, 1) their positions, i.e., *adjacency*, and 2) orientations, i.e., *surface normal*. The merging is allowed, if and only if two 3D plane-roof polygons are I) adjacent, and II) have approximately the same orientations in space (coplanar). The first criterion is derived from the PAR computed above and it is checked first. The second criterion can be mathematically introduced as the equation (5.1). It physically states that the angle between the surface normal vectors $\vec{n_1}(a_1, b_1, c_1)$, and $\vec{n_2}(a_2, b_2, c_2)$, of the two respective adjacent polygons p_1, and p_2 must be less than a pre-specified threshold. If merging is permitted, then two polygons are merged into a new polygon, and the two old polygons are discarded.

$$\cos^{-1}(\frac{\vec{n_1} \cdot \vec{n_2}}{|\vec{n_1}| \cdot |\vec{n_2}|}) < t_{angle} \tag{5.1}$$

As rule of thumb, the threshold parameter t_{angle}, can be related to the geometric quality of the existing DSM, and estimated fitting error obtained during the back projection of the 2D polygon primitives into the 3D object space. Assuming that the additive noise (error) process is relatively stationary within every region of interest in DSM, the same strategy developed in section (4.5.1) for estimating the noise variance in the image can be utilized in order to tie the above algorithmic threshold into the amount of error in DSM.

Since the geometry and topology of the polygons might change after each merging process, the algorithm works in an iterative procedure and stops when no further merging is possible. For every region of interest, the algorithm selects the largest 3D polygon p_1, which is probably the most significant geometric primitive of the associated roof structure, and checks the compatibility requirement (equation 5.1) with respect to its adjacent polygons. It starts with the smallest adjacent polygon p_2, excluding the 'contained-in' polygons. If merging is permitted, the *weighted-union* of the two polygons is computed and replaces the larger polygon. The smaller polygon p_2, is eliminated from the list and the PAR of the new polygon p_1, is updated accordingly. The weighted-union means, that the surface normal vector $\vec{n_1}$, of the new polygon is computed based on the weighted mean of the two old normal vectors $\vec{n_1}$, and $\vec{n_2}$, and is formulated as:

$$\vec{n_1}^{new} = \frac{s_1\vec{n_1} + s_2\vec{n_2}}{s_1 + s_2} \tag{5.2}$$

where s_1, and s_2 are the sizes of polygons p_1, and p_2 respectively. In this manner, the larger polygon has more contribution into the computation of the new polygon. This process is repeated between the new polygon p_1, and all its adjacent polygons based on the updated PAR, until no further merging is possible. The second largest polygon in the list is selected next, and the above procedure is repeated. The whole merging algorithm stops when all the current polygons within the region of interest are evaluated for the compatibility criteria. Note that the 'contained-in' polygons are dealt with differently -in fact, when all the adjacent polygons are processed, and the PAR updated. The merging process is performed for every 'contained-in' polygon with respect to the polygon, which contains it. In these cases one more compatibility requirement must be met as well. A 'contained-in' polygon is permitted to merge, if its deviation, with respect to its altitude, from its 'contains' polygon is less than some threshold. That means if the two polygons have the same orientation, but the 'contained-in' polygon is higher e.g., $0.5m$, then it is part of a microstructure in top of the roof and is not allowed to merge.

Theoretically, the whole merging process can be performed in a single sweep using only one strictly angle threshold t_{angle}. But in practice, in order to avoid early errors of incorrectly merging adjacent but distinct planar polygons, it is realized that the better results are achieved, if the merging process is performed more than once, e.g., 3 sweeps, starting with the small value of the angle threshold and increasing its value in each subsequent sweep. In addition, as it is discussed earlier very small regions may result during segmentation. Some of these regions are not parts of the physical roof structures, thus they neither satisfy the compatibility requirements to be merged nor are large enough to stand individually as the significant parts of a building roof and should be eliminated from the list of roof primitives. Therefore, a threshold is set on the size of polygon primitives, after the merging operation. For very small polygons a priori knowledge is available that they are too small to exist on their own. This a priori knowledge results from the problem description.

The results of the merging algorithm described in this section are shown in figure (5.7). The primary fragmented plane-roof polygons of a gable (5.7-a), and a complex roof structure building (5.7-c), are employed. The result of the proposed merging algorithm in figure (5.7-b), shows that the compatible adjacent polygons are correctly merged into the larger polygonal primitives, which have the meaningful correspondence to the major parts of the roof structures. Figure (5.7-d) indicates that the irrelevant noisy small polygons are also filtered out during the merging process. In addition, the algorithm is capable of preserving the small 'contained-in' polygons

corresponding to the dormer window structure on top of the roof building. Although the shape, and somehow the geometry of these micro elements are not accurately defined or some of them are totally eliminated or merged into the larger polygons, their presence, however, at the time being significantly improves the results of the subsequent reconstruction processes. Moreover, it reveals that the reconstruction of these types of microstructures is feasible with the improvement of the available imaging sensor and utilizing the high accurate DSM.

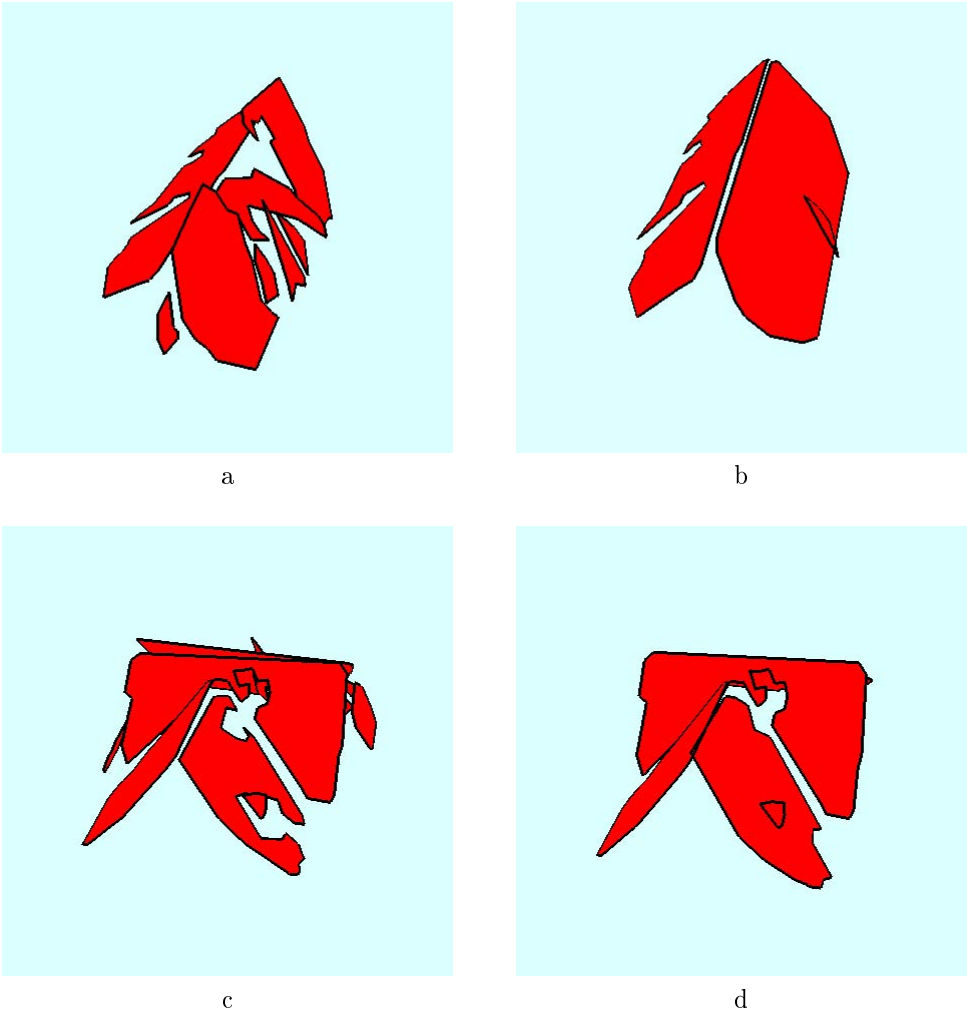


Figure 5.7: Merging adjacent 3D plane-roof polygons, a) and c) the 3D perspective view of the primary 3D plane-roof polygons of a gable and a complex roof structure building respectively, b) and d) the corresponding 3D plane-roof polygons after merging operation

5.5 POLY-MODELER: Generic Polyhedral-Like Model Generator

This section introduces a novel method for the geometric reconstruction of a plane-face solid object, commonly called *polyhedral*. Polyhedral is an arrangement of polygons such that two and only two polygons meet at an edge. Simple polyhedral refers to all polyhedra that can be continuously deformed into a sphere and are convex. The term convex applies to every polyhedral that lies entirely on one side of each of its polygonal faces. This is our sign to investigate *non-simple polyhedra*. They are topologically equivalent to any complex solid that may have holes in it and/or is concave and are, therefore, of direct use to us in geometric reconstruction of complex plane-roof buildings. There are several approaches to represent complex solid models. The two common methods are *constructive solid geometry (CSG)*, and *boundary representation (b-rep)*. The CSG scheme defines complex solids as Boolean combination of simpler solids. The complete representation is sometimes referred to as a CSG tree, because it uses a binary tree whose terminal nodes are simple solids and whose non-terminal nodes are so-called regularized Boolean combining operations. The b-rep describes the faces, edges, and vertices of the

boundary of the solid. This description itself has two forms; a topological representation of the connectivity of the boundary elements, and numerical data describing the shape geometry and position of these elements. If object perception is primarily dependent on surface perception, it is a natural choice for an automated vision system to use a boundary-based technique to represent the object of interest. This is why we have selected a b-rep method for modeling and representing the generic building models and their primitive elements. In this study b-rep scheme describes a generic building model as the union of very general faces embedded in unbounded plane-surfaces, where the building edges are defined by the intersections of these surfaces. Such generic models can be constructed directly by assembling and intersecting appropriate surfaces. An algorithm called POLY-MODELER performs the reconstruction. The algorithm determines where component faces are extended or truncated and new edges and vertices are created or deleted. When boundary elements overlap or coincide, the algorithm merges them into a single element and thus maintains a consistent, non-redundant data structure representing building model boundary. New edges are created where adjacent faces (polygons), intersect. The POLY-MODELER finds these intersections and then determines by *point membership classification*, which segments of the intersection are actual edges of the model. It should be noticed that the proposed method is designed to reconstruct any polyhedral-like object model but the main intention is the 3D reconstruction of generic plane-roof buildings. In this application we only concentrate on the description of the shape and form of the roof, once the complete building roof is modeled, the fictitious vertical walls are incorporated to the model to generate a complete solid building model.

Modeling complex objects such as buildings requires considerable attention to their topology. We must understand how simple elements are connected to form the complex model and how its topology is preserved when subjected to a variety of transformations. Topological properties are not metrical, but concern such things as connectivity and dimensional continuity. In the previous sections we briefly discussed processes of construction of the primitive 3D plane-roof polygons (faces) and how to compute their adjacency relationships (PAR). All of these form the foundation on which we build the generic building model, which is the main contribution of this chapter.

5.5.1 Basic Notation

To proceed, first certain concepts and notation schemes that are used to describe and express some of the processes and components involved in POLY-MODELER are discussed. The basic idea in the roof modeling process is to combine adjacent 3D plane-roof polygons, shortly *3D-poly*, to construct the roof structure. A stitching operation would be a logical way to glue the adjacent 3D-polys along their common edges. The 3D intersections play a prominent and manifest role in this operation. Typically when two 3D-polys must share a common edge, each of them is arbitrarily considered as an unbounded 3D plane in space. They intersect and geometrically the line of intersection is then determined.

Figure (5.9-a) shows a perspective view of two adjacent 3D plane-roof polygons p_1 and p_2 along their computed line of intersection l_1. The top view of the corresponding polygons is illustrated in figure (5.9-b). To complete the stitching operation, depending on the shape and geometry of the 3D-polys, the unwanted part of the 3D-polys are trimmed along the intersection and/or the boundary of 3D-polys are extended until they reach their physical limit, which is the line of intersection. Figures (5.10-a), depicts a perspective view of the final result of the stitching operation. The grey part of two polygons indicates the extension of 3D-poly p_1 and p_2. The part of 3D-poly p_2 (see figure 5.10-b), which is bounded with the dotted line indicates the trimmed part of this polygon.

Before moving to the mathematical aspect of this operation and in general the proposed 3D reconstruction method, the following simple definitions that are essential concepts and tools for solving the task are introduced.

Definition:

- *Coordinate system*; the right-handed Cartesian coordinate system is assumed unless noted otherwise.
- *Polygon parameterization*; the 3D plane-roof polygons are stored as a single-sided face in POLY-MODELER. A single-sided face means that points on one side are considered to be inside, in this context this side is called the *sense* of the 3D-poly, and points on the other side are considered to be on the outside. Therefore a consistent ordering of the polygon vertices is important, because it represents the sense of the 3D-poly. The vertices sequentially are numbered in clockwise direction. In this way the surface normal at any point always points toward the interior of the reconstructed building model as it is shown in figure (5.8). The resulting 3D-poly is called a parameterized polygon.

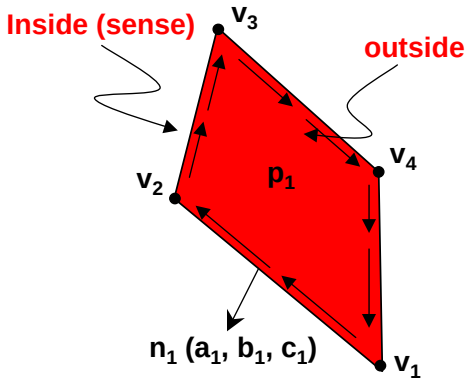


Figure 5.8: Polygon parameterization.

- *Symmetric polygons*; let $\vec{n}_1(a_1, b_1, c_1)$ and $\vec{n}_2(a_2, b_2, c_2)$ indicate the surface normal vectors of the two adjacent 3D-polys p_1 and p_2 respectively. In this context, p_1 and p_2 are *symmetric*, if they satisfy all the following conditions:

$$\begin{aligned} a_1 &= -a_2 + \delta_a; \\ b_1 &= -b_2 + \delta_b; \\ c_1 &= c_2 + \delta_c; \end{aligned} \tag{5.3}$$

The parameters δ_a, δ_b, and δ_c represent the sum of small deviations of the surface normal vectors $\vec{n}_1$, and $\vec{n}_2$, with respect to their original values. The values of these tolerances are directly related to the quality of the derived DSM, which is used to recover the orientation of the 3D-polys in object space.

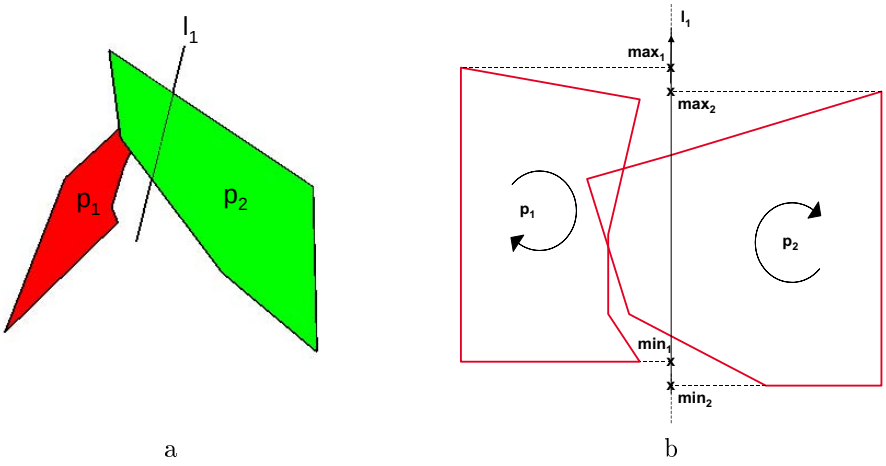


Figure 5.9: 3D intersection of adjacent polygons.

- *Line of intersection*; geometrically the 3D intersection of two non-coplanar unbounded adjacent 3D plane-faces (polygons) p_1, and p_2, in E^3 determines an unbounded straight line l_1, (figure 5.9-a). The computed line corresponds to an orientation-edge in the roof structure, such as ridge or saddle edge of the roof. The intersection line has also an interesting property; just as a point separates a line into two parts, so does a

line separate an unbounded plane into two half-planes in E^2. In addition if an arbitrarily unbounded 3D-plane, e.g., a vertical plane, passes through the line, it divides the space into two half-spaces in E^3. This is the milestone of our approach, *the 3D reconstruction of plane-face solid models is based on simultaneous intersection of half-planes.*

- *Point of intersection*; the 3D intersection of three non-coplanar unbounded adjacent 3D plane-faces p_1, p_2, and p_3, in E^3 determines an intersection point p_{int}, as it is shown in figure (5.11). These points are the most significant and accurate geometric primitives of the reconstructed model. Besides of their structural characteristic, they also provide significant semantic information concerning extendibility of the incorporated 3D-polys with respect to each other. The concept will be elaborated mathematically in the follow-up section. These points enter to the reconstruction process as new vertices. The membership property of the intersection point p_{int}, as it is a new vertex of three adjacent polygons is stored as a connectivity relationship in POLY-MODELER.

- *Minimum and maximum points*; they, e.g. min_1, max_1, indicate the bounding extension of 3D-poly p_1, on the line of intersection l_1, with respect to its adjacent 3D-poly, here p_2 (see figure 5.9-b). In fact, physically, they are the intersection points between bounding edges of the polygon with the line of intersection. Accordingly in a multiple case, that is when a 3D-poly has more than one adjacent polygons, the intersection operation determines multiple pairs of such points which are considered as model point candidates during point membership classification in POLY-MODELER (see the small triangles in figure 5.10-b). The terms minimum and maximum express the order of the points in the direction of the intersection line. It should be noticed that when minimum and/or maximum points of a 3D-poly such as max_2 of p_2 lie between the minimum and maximum points of its adjacent polygon, then we also consider these points as new candidate vertices for its adjacent polygon. Moreover, if the minimum point of a 3D-poly is located at a close neighborhood of the maximum point of its adjacent polygons, then both extreme points are replaced by their average mean point p_{mean}. The membership property of p_{mean} which is the member of both 3D-polys also stores as a connectivity relationship in POLY-MODELER.

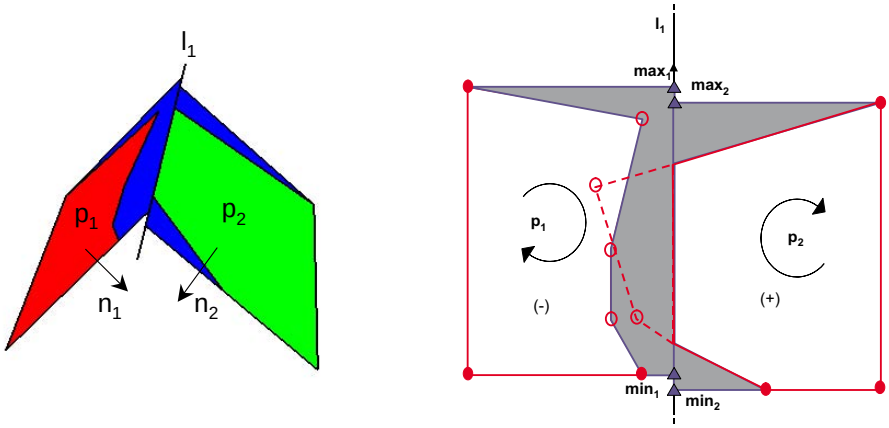


Figure 5.10: Stitching operation of adjacent polygons

- *Polygon status and extendibility*; in POLY-MODELER we use the term extendibility either a polygon will be extended or trimmed during stitching operation. Based on this definition, after computation of intersection line, the POLY-MODELER labels the *status* of every 3D-poly as positive (forward), or negative (reverse), or zero. The positive status means that the polygon extension must be carried out in a clockwise direction starting from the minimum point towards the maximum point i.e., polygon p_2 in figure (5.10-b). In the contrary, the negative status means that the polygon extension must be carried out in a clockwise direction but starting from maximum point toward minimum i.e., polygon p_1 in figure (5.10-b). The flag zero means that the 3D-poly is not extendible with respect to its adjacent polygon, such as a 3D-poly which will not extend with respect to its 'contained-in' polygon. The polygon status is a tool to determine which part of the 3D-poly (half-plane) is inside the building model. This is why polygon extendibility is a mutual relationship. That means if one of the polygon is labeled as positive, the respective adjacent

3D-poly must be a negative polygon. This is a sign for classification of polygon points as active or inactive points, which will be discussed next. It should be noticed that in a multiple case, a 3D-poly might have as many status labels as it has adjacent polygons which are not necessarily the same.

- *Redundant vertices*; in a compact and reliable geometric modeling process, it is desired to exclude the redundant geometric primitives such as vertices, or edge segments. In POLY-MODELER we classify polygon points into two classes. The first class called *active vertices*, black circles in figure (5.10-b), are those vertices which are considered as candidate points in the reconstruction procedure, whereas *inactive vertices*, empty circles are points that must be shifted on the intersection line during stitching operation and are located between two extreme points. Geometrically these points are collinear, so that they are considered as redundant points and are excluded from the proceeding process. As we have already mentioned the status of each polygon initializes the task. Consider walking around 3D-poly p_2, in a clockwise direction, based on its status, here positive, we flag all the vertices as inactive when moving from min_2, toward max_2, point. In the contrary when we walk around p_1, again in a consistent clockwise direction, we consider all the points which are located between max_1, and min_1, points as inactive or redundant points when walking from max_1, toward min_1.

We have defined certain notations and conventions that are used in the POLY-MODELER algorithm. We have also discussed some of the operations and classification aspect that are involved in 3D intersection process. Thus, we have most of the ingredients for the next section, which introduces the mathematical aspects of how POLY-MODELER works.

5.5.2 Mathematical Concept and Methodology

The objective of a b-rep modeler such as POLY-MODELER is to build a complete representation of a solid as an organized collection of surfaces. In general a b-rep model stores the numerical data of the surface geometry on which the face lies, the curve geometry on which the edge lies and which bounds the face, and the point geometry of the vertices. In fact, the POLY-MODELER is obviously a special case of boundary representation when curved surfaces and edges are approximated by planes and straight lines respectively. It should be noticed that there is a minor deviation in how POLY-MODELER works in special application of generic building reconstruction from its original design strategy. Since the vertical walls of the buildings are perpendicular to the ground and are simply reconstructed based on the outline of the building roof, the POLY-MODELER only concentrates on reconstruction of roof structure. However, in the final step the fictitious vertical walls are added to complete the reconstruction of a plane-face solid building model. We stated in the earlier section that the objective of our work is to reconstruct non-simple generic polyhedral models. In other words, the method is designed to handle any type of polygonal faces, such as *concave polygons*. In this context, we distinguish two classes of concave 3D-polys. The first class is when the concave part of the 3D-poly lies on the outline of the roof, similar to bounding edges of the 3D plane-roof polygons in figure (5.7-b), which are incorrectly defined during segmentation process, due to the presence of noise, shadow, and microstructure on top of the building roof. Reconstruction of this type of 3D-poly is a trivial process. Its concave parts are recovered in a model-based approach based on a fitting algorithm during consistency verification of coarse building hypotheses against 2D image primitives, and is discussed in chapter 6. The second class refers to those polygons for which their concave parts are geometrically constructed by intersection of three adjacent 3D-polys (see 3D-poly p_1 in figure 5.11-b).

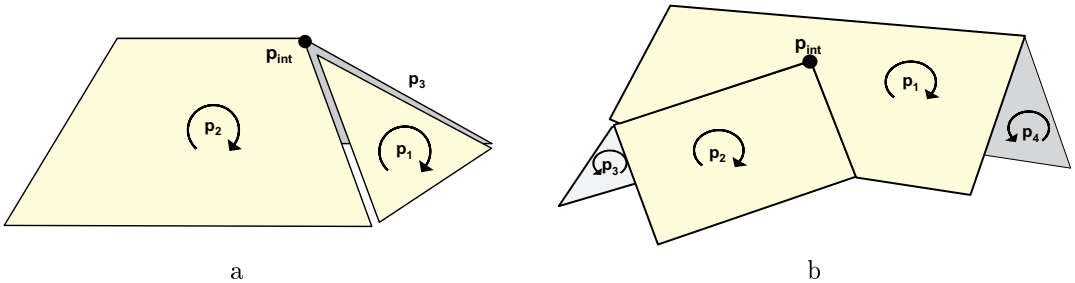


Figure 5.11: Analysis of the point of intersection between three adjacent 3D-polys gives an indication of polygon convexity, a) 3D-poly p_1, is a convex 3D-polygons, b) 3D-poly p_1, is a concave 3D-polygons

In fact, this is the most ambiguous part of a generic data-driven reconstruction procedure. A failure in detecting the concave polygons during intersection operation may cause to cut off accidentally the active parts of a roof structure. In order to correctly flag the status of such a polygon, we developed a new approach for the detection of concave 3D-poly and its extendibility status, based on the presence of an intersection point p_{int}.

Let $\vec{n}_1$, $\vec{n}_2$, and $\vec{n}_3$ indicate the *unit surface normal vectors* of the 3D-polys $p1$, $p2$, and $p3$ in any right-handed Cartesian coordinate system respectively. Furthermore assume that p_2 and p_3 are symmetric and both are asymmetric respect to p_1 (figure 5.11). According to building construction rules, we may now assert the following axiom:

Extendibilty axiom: *The 3D-poly p_1 is extendible with respect to its adjacent polygons p_2, and p_3, if and only if it has a sense-to-sense relationship with p_2 and p_3. In other words, when the normal vectors of all three polygons are convergent inside the model.*

To realize the concept, at presence of every intersection point, p_{int}, an extendibility analysis is performed based upon computation of a *scalar triple product* of the three unit surface normal vectors in the following order:

$$v = \vec{n}_1 \cdot (\vec{n}_2 \times \vec{n}_3). \tag{5.4}$$

The absolute value of the scalar triple product v, has a simple geometric interpretation. It is equal to the volume of a parallelepiped with $\vec{n}_1$, $\vec{n}_2$, and $\vec{n}_3$ as adjacent edges. We should keep in mind that in this analysis, the order of the product is significant. It is always *the scalar product of the asymmetric normal vector n_1, by the result of the cross product of symmetric normal vectors in a consistent clockwise direction* $(\vec{n}_2 \times \vec{n}_3)$. Since the surface normal vectors are unit vectors, the maximum value of the scalar triple product is $v = \pm 1$. The minus sign indicates a transition from a right-handed to a left-handed system (or conversely) which is a signal for further analysis. In general, various geometric configurations of adjacent polygons produce different results as follows:

1. $v = \pm 1$; indicates that p_1, p_2, and p_3 are perpendicular. In this case at least one of the 3D-polys, p_1, is a vertical wall. In practice, this case will not happen, as we do not consider vertical walls during the reconstruction stage. However, in general, we will extend p_1 with respect to p_2 and p_3 based on a simple 3D intersection operation.

2. $v = 0$; shows that at least two of the 3D-polys are coplanar. This is not also a practical case, because the coplanar 3D-polys are already merged in preceding processes. We assign polygon status of p_1 respect to p_2 and p_3 to zero.

3. $0 < v < 1$; indicates that all three polygons are *sense-to-sense*, as it is shown in figure (5.11-a). The 3D-poly p_1 will extend in a simple manner based on normal 3D intersection operation.

4. $-1 < v < 0$; indicates that the scalar triple product, produces a left-handed Cartesian coordinated system. Geometrically, it means that surface normal $\vec{n}_1$ is not pointing towards the other two vectors. It also indicates that 3D-poly p_1 is a concave polygon which is constructed because of the presence of intersection point p_{int} (see 3D-poly p_1, in figure 5.11-b). Thus p_1 is not extendible with respect to p_2 and p_3. However, to recover the geometric shape of the p_1, we do further analysis as follows:

Figure (5.12) shows four possible geometric configurations between an asymmetric 3D-poly p_1 and two symmetric 3D-polys p_2, and p_3, along their intersection point p_{int}, when the result of the scalar triple product is a negative value $(-1 < v < 0)$.

The signs '+', and '-' indicate different status of p_1, and its respective polygons. The small triangles show the new minimum and maximum points of p_1, on the line of intersections. For example, point max_{1-2} of *'case I'* represents the maximum point of p_1, which is constructed by the 3D intersection of p_1 and p_2. We discussed previously that based on any 3D intersection between two adjacent polygons we insert two new polygon vertices, minimum and maximum as candidate points, into the stitching operation of that polygon. Therefore in an ordinary intersection process such as figure (5.11-a), we consider 5 candidate points ($= min_{1-2}$, max_{1-2}, min_{1-3}, max_{1-3}, and p_{int}) as new vertices during reconstruction of p_1. While in the special case of a concave polygon, only one point from each extreme point groups is considered as a new polygon vertex. In other words for 3D-poly p_1, in the *'case I'*, we only consider three ordered points such as $(max_{1-2} \rightarrow p_{int} \rightarrow min_{1-3})$ as

new vertices during the reconstruction process. This selection is explained as follows. The status of 3D-poly p_1, with respect to 3D-poly p_2 is negative, therefore parameterization start from max_{1-2} in a clockwise direction towards intersection point p_{int}, disregarding the min_{1-2} point, as it is an inactive point. Contribution of this point into the roof reconstruction process causes that the valid part of the 3D-poly p_1, is trimmed off. In a similar manner, continuing the parameterization with respect to 3D-poly p_3, the status is negative, thus starting from max_{1-3} point, which is replaced by intersection point p_{int}, moving in a clockwise direction, towards min_{1-3}. Accordingly, other cases are handled as it is shown. In addition, we change status of p_1 to zero. This preserves its extension in a subsequent process respect to p_2, and p_3. Now that all essential analyses are applied, we are ready to perform the final reconstruction operation.

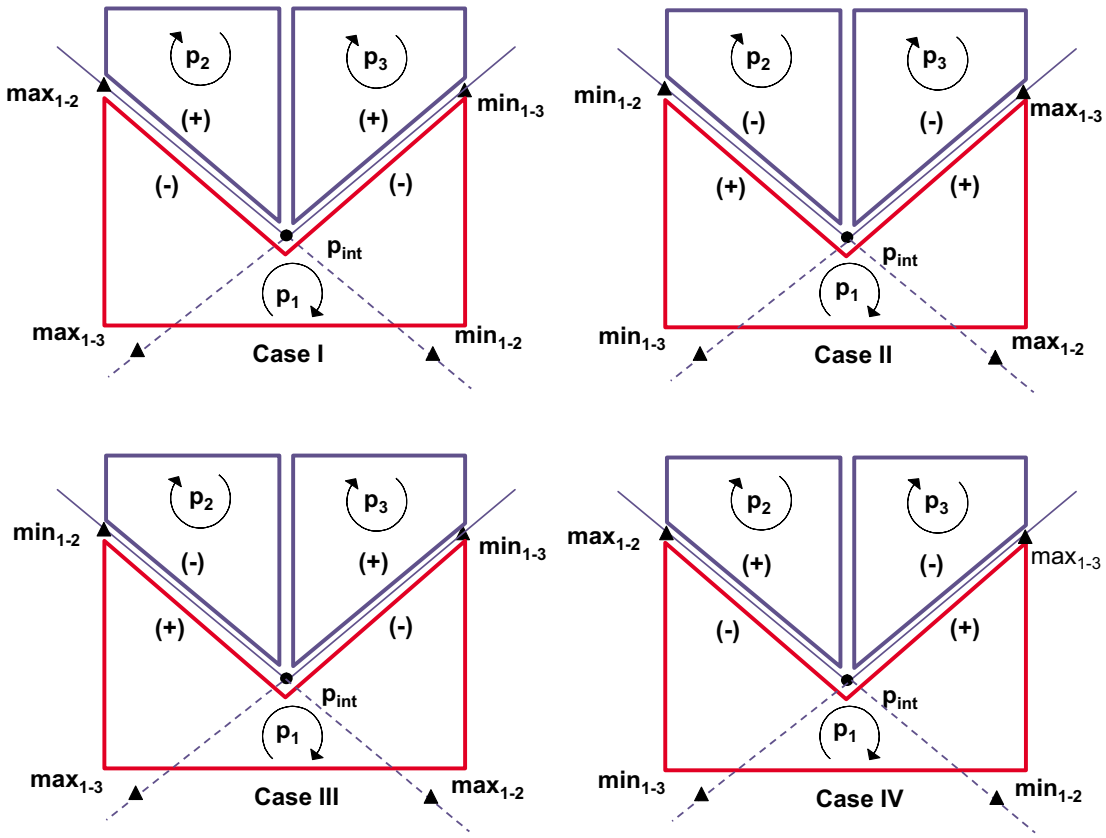


Figure 5.12: Analysis of concave polygons

Let $R^{2,3}$ indicate a two dimensional region, e.g., a plane-surface, which is located in space E^3. We will write

$$R^{2,3} = [R_i^{2,3}, R_b^{1,3}] \tag{5.5}$$

where $R_i^{2,3}$ is a set of 2D regions in the interior of the region $R^{2,3}$, and $R_b^{1,3}$, indicates a set of 1D regions, e.g., straight-edge, on the boundary of the $R^{2,3}$. Furthermore any point in space has one and only one of the following three properties with respect to any region $R^{m,n}$ (Mortenson 1997).

1. It is inside the region; that is, it is a member of the set $R_i^{m,n}$.
2. It is on the boundary of the region; that is, it is a member of the set $R_b^{m-1,n}$.
3. It is outside, not a member of the set $R^{m,n}$.

Thus, for a homogeneous surface in E^3, the explicit definition of the $R_b^{1,3}$ of the surface is necessary and sufficient for the definition of the surface ($R^{2,3}$). $R_b^{1,3}$ is the outline or boundary of the surface, where points on the inside ($R_i^{2,3}$) are implied by $R_b^{1,3}$. We will now take a more general approach for reconstruction of polyhedral-like object models, such as complex building models. Taking a collection of planar pieces and gluing them together along their edges, creating piecewise flat-surfaces forms a building roof structure. Any surface formed in this

way will obviously be flat everywhere except possibly along the edges where the pieces are glued together. The crucial step is in deciding how to reconstruct the plane-faces based on their respective bounding edges. We compute the edges of these faces based on the 3D intersection approach. The intersection line defines the valid half-plane in E^2, (for simplicity of the proceeding computations, we consider an orthogonal projection of all 3D primitives such as plane-polygons, intersection lines and intersection points in a two dimensional space E^2). Thus, disregarding the part of the faces that are not embedded in the specified half-plane. This is done based on a point membership classification method that excludes candidate model points, which are not inside or on the boundary of the object model. The ultimate objective is the computation of the common intersection of n half-planes, where n is the number of non-zero polygon status. This is realized by a simultaneous solution of n linear inequalities in the forms of (5.6), and (5.7), which is in fact a linear programming problem, where the objective function to be maximized is the area of the respective 3D-poly, while at the same time a set of conditions such as the following inequalities should be satisfied (Fryer 1978, Best & Ritter 1985, Fritsch 1985), .

$$f_i(x,y) \leq 0; \quad \forall \quad i = 1,2,...,k \tag{5.6}$$

$$f_j(x,y) \geq 0; \quad \forall \quad j = 1,2,...,l \tag{5.7}$$

$$n = k + l$$

where $f(x,y)$ is the equation of the intersection line in E_2. The inequality (5.6) stands when the polygon status with respect to its adjacent polygon is positive, while the inequality (5.7) is given for negative status. In practice the desired solution, *feasible region*, is achieved by simply testing all the candidate points p_c, including the active polygon vertices and new vertices such as intersection points, against the set of inequalities defining the half-planes. As we proceed through an ordered list of these inequalities, we update a flag on p_c. As long as p_c, satisfies each successive inequality constraint, it is flagged as active points. If the p_c, fails any test, the loop terminates, and we flag the p_c as inactive point, that means it is outside the roof structure. The final shape and description of 3D plane-faces are constructed by parameterized concatenations of the remaining active points in 3D space. It should be noticed that during construction of concave polygons, new intersection points and their associated extreme points will not be tested against inequality constraints. Instead they will contribute in the reconstruction process without any conditions.

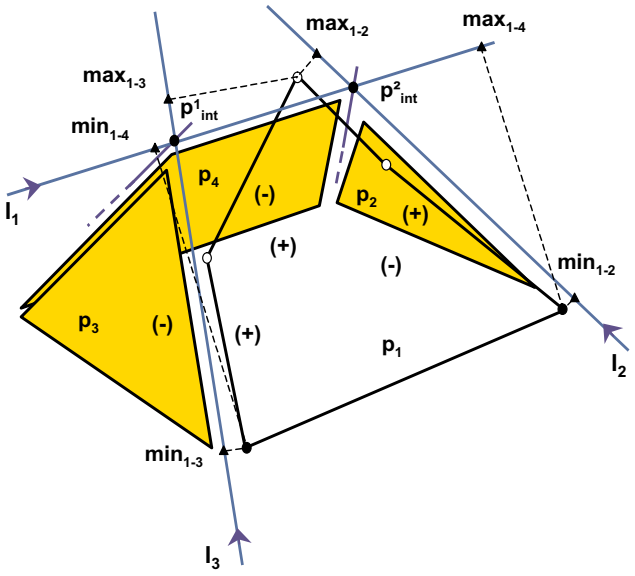


Figure 5.13: An example of the roof modeling based on the 3D intersection of the adjacent plane-faces

Figure (5.13), illustrates an example of roof modeling based on the proposed method. Let 3D-polys p_1, p_2, p_3, and p_4 indicate the plane-surfaces of a building roof structure. Furthermore assume that 3D straight lines l_1, l_2, and l_3 are computed based on the intersection of 3D-poly p_1, with its adjacent polygons p_4, p_2, and p_3, respectively. The associated extreme points min_{1-2}, max_{1-2}, min_{1-3}, max_{1-3}, min_{1-4}, max_{1-4}, and the 3D intersection point p^1_{int}, and p^2_{int}, are also defined accordingly and are introduced into the POLY-MODELER

as the new candidate vertices of the 3D-poly p_1. For the simplicity, only the associated geometric primitives of the polygon p_1 are shown in the figure. To proceed, first all the 3D primitives are transferred into the 2D space based on an orthogonal projection. Figure (5.14) shows a top view of the corresponding elements in 2D space. Next, the inequalities conditions are defined based on the equations of the intersection lines and the status labels of polygon p_1 with respect to its adjacent polygons. The intersection lines l_1, and l_3 define two inequality conditions of the form (5.6). In contrary, the intersection line l_2 introduces an inequality condition of the form (5.7), as the status of polygon p_1 with respect to polygon p_2, is negative. As it was discussed previously, every line divides the space E^2 into two half-planes and the polygon status consequently define the valid half-plane. The simultaneous intersection of these half-planes determines the valid or feasible region of the 3D-poly p_1, which is indicated as gray area in the figure (5.14).

The POLY-MODELER defines the extension of the feasible region, simply by flaging all the polygon vertices v_i, and the new candidates as either active, if they satisfy all the inequality conditions, or inactive, if they fail. For example, the extreme point max_{1-3} is failed to satisfy the conditions introduced by line l_1, it is located in the invalid (hachured area in figure 5.14) half-plane, thus it is flagged as an inactive point. The vertices v_4, and v_2 are also flagged as inactive points. In fact, these vertices labeled as inactive points in previous processes before assessing upon the inequality conditions, because they lie on the bounding edges of the polygon p_1, thus geometrically are redundant points, and should be eliminated in order to have a consistent and compact b-rep model of the roof structure.

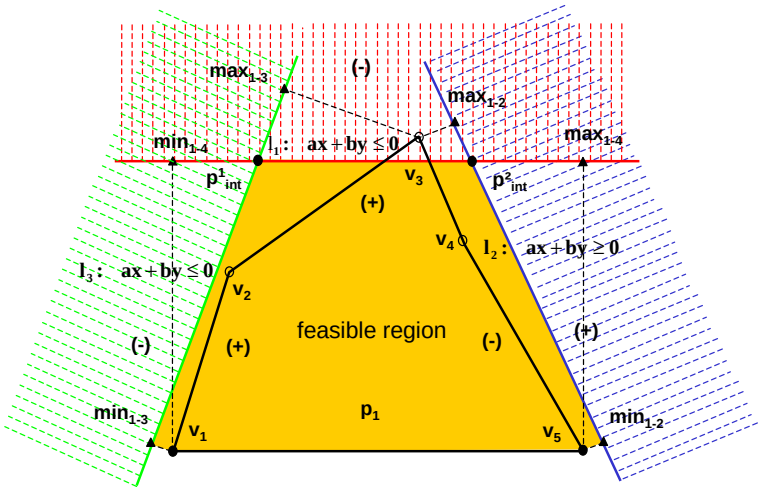


Figure 5.14: Determination of the feasible region (extension of the plane-surfaces of the roof model) by simultaneous solution of a set of n linear inequality conditions.

The final shape and extension of the 3D-poly p_1 is determined by parameterized concatenations of the remaining active points in 3D space, that is ($v_1 \rightarrow min_{1-3} \rightarrow p^1_{int} \rightarrow p^2_{int} \rightarrow min_{1-2} \rightarrow v_5$). Similarly, the extensions and the geometric descriptions of other 3D-polys of the roof are determined, including the membership properties and connectivity of the common points and edges between adjacent polygons. These elements are stored as geometric primitives of the roof structure and represented as b-rep model of the building roof.

5.6 Experiments and Results

Figure (5.15), illustrates result of the proposed method and the performance of the POLY-MODELER algorithm. Three buildings with different roof structures are selected. The figures in first column, (5.15-a), (5.15-d), and (5.15-g), depict the 3D plane-roof polygons of a gable, hipped-gable, and a complex roof structure before extension. The second column illustrates the 3D perspective view of the corresponding roof models after reconstruction in object space (5.15-b), (5.15-e), and (5.15-h). Third column shows a top view of the reconstructed roof models overlaid on corresponding building roof. For each building candidate or region of interest, the geometric descriptions of all the associated polygonal primitives of the roof and their adjacency relationships computed previously by PAR are introduced into the POLY-MODELER as initial parameters. The POLY-MODELER, then generates a very dense internal pointing data structure to keep track of the changes of all the

primitives during reconstruction process. The POLY-MODELER models the roof structure in a generic manner purely based on the 3D intersection of adjacent polygons, without any a priori information concerning the roof type or imposing any external constraints. That means the same procedure is applied to any type of roof with any number of polygonal primitives and any complexity. In this example, the selected complex building has interesting properties, it simply invalids many of the constraints applied in *specific model-based* approaches. For example, constraints on orthogonality of the building outline cause that reported methods fail to reconstruct a correct and accurate model of this type of generic building.

As discussed earlier, the result of the reconstruction process is highly related to the estimated parameters of the 3D plane-roof polygons. A failure in correctly recovering the surface normal of the 3D-poly will cause an unexpected result leading to a partially or completely wrong building description. This is why the quality of utilized DSM is of high importance in our approach. The results indicate that the recognition of microstructure on top of the building roof such as a dormer window is also possible. Nevertheless, in order to be able to estimate the correct pose of these microstructures on top of the roof and geometrically describes their shapes, a very dense and high accurate DSM is required. Furthermore, owing to the geometrical reconstruction of roof structure, positional accuracy of roof elements such as orientation edges and intersection points are very high. However, due to a misinterpretation of surface normal of polygonal primitives, we may have some discrepancies in the form of displacement or rotation from the real positions of these elements, see the intersection point of the three planar faces of the complex roof structure in figure (5.15-h).

In addition, due to the nature of the region-growing type segmentation, the quality of roof outline of the building model is still poor (see figure 5.15-b). The region growing type segmentation methods are unable to accurately localize the bounding edges, and the shape of the region boundaries normally reflects the search strategy than the true shape of the region. This is the reason why the generated hypothesis model is called *coarse building model*. To improve the quality of the generated model, the geometric and topological information provided by the coarse model is incorporated into the hypothesis model verification process. A fine building model is obtained in an iterative top-down estimation process called *Feature Based Model Verification (FBMV)*. This is done by simultaneously fitting the 3D model primitives into the corresponding 2D image features where the geometrical and topological model information is integrated into the process as external and/or internal constraints. The detailed discussion of this process is the topic of the next chapter.

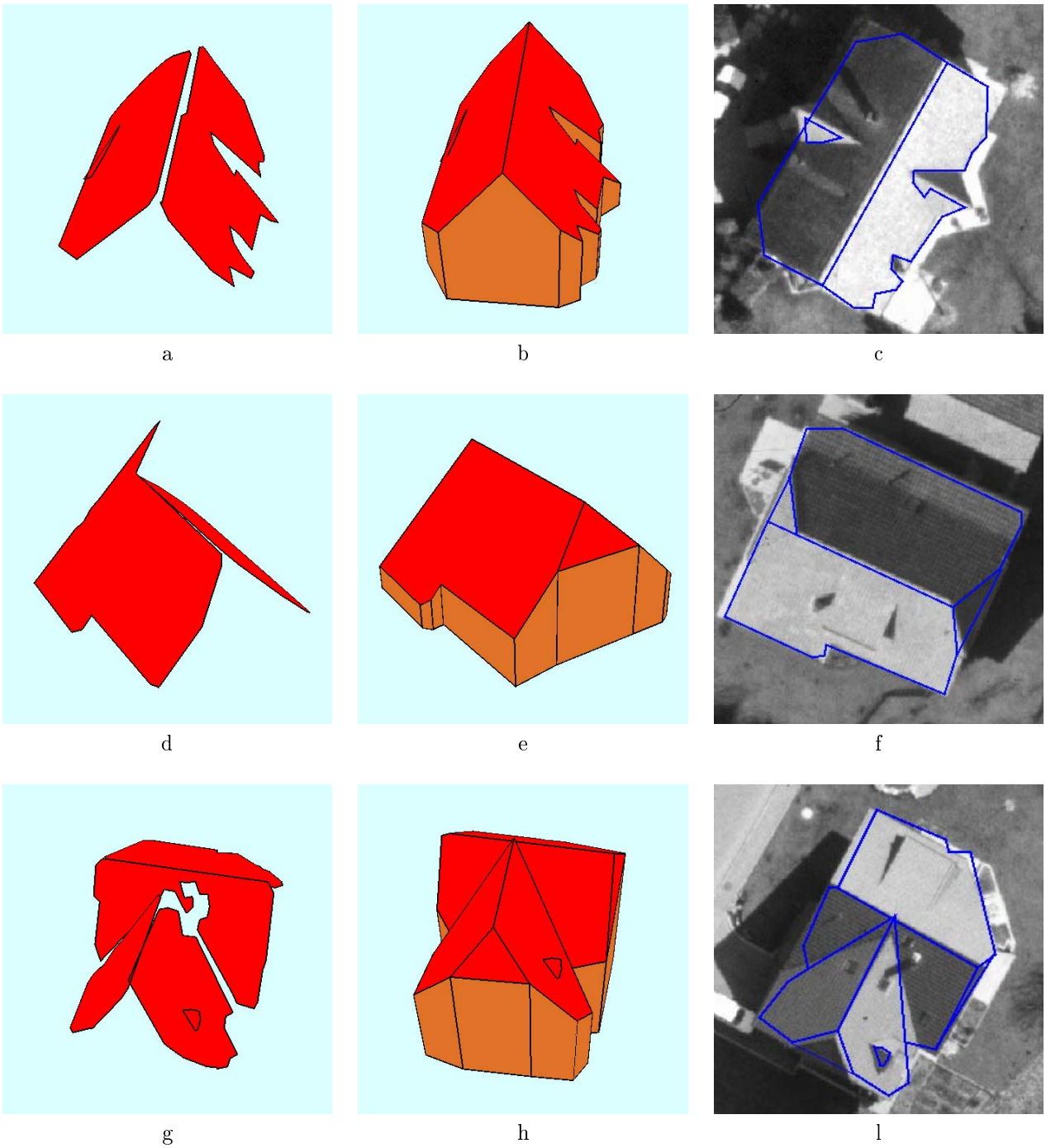


Figure 5.15: 3D building roof modeling, a), d), and g) extracted 3D plane-roof polygons of a gable, hipped-gable and a complex roof structure before roof modeling respectively, b), e), and h) 3D perspective view of the corresponding reconstructed 3D coarse building roof models, c), f), and l) reconstructed coarse building roof models overlaid on corresponding building roof structures in 2D image space.

Chapter 6

Feature Based Model Verification

6.1 Introduction

This chapter introduces an automated method called *Feature Based Model Verification* (FBMV), for modification and verification of the reconstructed generic polyhedral-like building model by back projecting the 3D model into the corresponding 2D images taken from different viewpoints. Treating the hypothesis model as evidence leads to a set of confidence intervals in image space that can be used as a search space to find the corresponding 2D image primitives and performing a consistency verification of the reconstructed coarse model. Theoretically, in stereo image analysis systems it is possible to solve the unknown parameters of a 3D model by matching the corresponding 2D images. However, in practice, the reliability and accuracy of the parameter determination can be substantially improved by simultaneously fitting the model into the images taken from more than two viewpoints. The methods presented here can be used in either situation.

The other important aspect of the FBMV method is the ability to solve the model parameters by simultaneously fitting all the geometric primitives of the 3D model into all the homologous 2D image features. Taking into account the external and internal geometric and topologic properties of the model structures and imaging process as constraints during parameter estimation. This is important because it allows the earlier initial matches or the partial matches between the 3D model primitives and 2D image features force the location of other structural elements of the model. Thereby new matches are generated that can be used to verify or reject the initial estimated model parameters.

The other important aspect of the FBMV method is the ability to solve the model parameters by simultaneously fitting all the geometric primitives of the 3D model into all the homologous 2D image features, taking into account the external and internal geometric and topologic properties of the model structure and imaging process as constraints during parameter estimation. This is important because it allows the earlier initial matches or the partial matches between the 3D model primitives and 2D image features force the location of other structural elements of the model, thereby new matches are generated that can be used to verify or reject the initial estimated model parameters.

The problem considered in this chapter is to determine the precise geometric descriptions of a polyhedral-like building model given matches between the model primitives and the image features. However, the proposed framework allows different non-polyhedral object models to be used. A consequent of disregarding this restriction is that the projected model edges are not necessarily straight edges and the model faces are not inevitably planar surfaces, thus the geometric routines should be adopted with different geometry.

A brief discussion of the general framework of FBMV method and its internal workflow is given first. The subsequent sections look inside the method and give detail discussions on the fundamental concept of FBMV, its formulation and its robustness. The evaluation of the proposed method, its performance and statistical analysis of the result obtained by some experimental test concludes this chapter.

6.2 Motivation

In the previous chapter we have presented a new method for reliable generation of a coarse polyhedral-like building model. The positional accuracy of the reconstructed roof elements such as ridgelines of the roof structure is highly related to the quality of the extracted 3D plane-roof polygons. Failure in correctly estimating the orientation of the 3D plane-roof polygons in object space causes displacement and rotation of the ridgelines with respect to their exact positions during the reconstruction process. In addition, due to the nature of region growing type segmentation discussed in chapter 4, the quality of the roof outline is poor. In fact, in real-world images, object boundaries cannot be detected solely on the basis of their photometry because of the presence of noise, occlusion and various photometric anomalies. Therefore, methods for finding boundaries based on purely local statistical criteria are tied to error, finding either too many or too few edges based on

arbitrary thresholds (Fua & Leclerc 1990). To supplement the weak and noisy local information of the images and probable misinterpretation of the orientation of the 3D plane-roof polygon, the geometric and topological information that the coarse object model can provide is incorporated into the chain of the reconstruction process. This information is introduced into the process of the object model verification based on a weighted least squares minimization process. A fine building model is obtained in an iterative, top-down, model-driven estimation process by simultaneously fitting the 3D model into the corresponding images where the geometrical and topological model information are integrated into the process as external and/or internal constraints during the estimation. The ability to apply such constraints is essential for the accurate modeling of complex objects. In particular, when dealing with a generic object model, it is crucial that the model elements are both accurate and consistent with each other. For example, individual components of a building can be modeled independently, but to ensure realism, one must guarantee that they touch each other in an architectural way. The estimation procedure yields a description of the building that simultaneously satisfies all the constraints within all the images. As a result, it allows us to perform a consistency check and refinement of the model across all the images. Moreover, the ability of the estimation method to fuse the information and impose the geometrical and topological constraints over all the images increases the accuracy and reliability of the reconstruction.

In the same line of major image matching techniques, i.e. feature-based (Förstner 1986), and area-based least squares matching (Förstner 1982, Ackermann 1984), the proposed verification process is called Feature Based Model Verification (FBMV). Similar to feature-based image matching techniques where a set of image-driven geometric features such as points, or edges are utilized in one image to be matched to the homologous features in corresponding images in order to, e.g. describe the surface geometry of the viewed scene. The FBMV uses model-driven geometric primitives to be matched to the respective homologous features in corresponding images taken from different viewpoints in order to verify the geometric description of the object model. In recent years, there has been a considerable increase in the number of publications on parameters solving for model-based vision, when most of the work aimed at parameters solving for rigid objects (Lowe 1991, Haala 1995, Fua 1996, Fischer et al. 1998, Brenner & Haala 1998*a*). An interesting similar work is reported by (Gruen & Li 1997). Their method is a semi-automatic approach for 3D extraction of linear features. In fact, this is an extension of a point-wise least squares template matching method (Gruen & Stallmann 1991, Baltsavias 1991), where a deformable contour model is used as a template instead of a square or rectangle which is generally used in conventional least squares matching techniques. However, in our study, their work is categorized as an area-based object extraction or alternatively *Area Based Model Verification* (ABMV).

6.3 Feature Based Model Verification

The FBMV is an iterative, multi-photo, multiple-dimensions, feature-based estimation process. It is designed to determine a reliable and accurate geometric description of the 3D structure elements of a reconstructed coarse solid model. It is a multi-photo approach because its formulation is independent of the sensor model. Thus, it is possible to introduce as many corresponding images taken from different viewpoints into the process. In this manner, the object model can be checked for consistency across all the images, thereby increasing the reliability of the reconstruction. It is considered multiple-dimensions, because a simple, but mathematically founded collinearity equation is used to precisely establish a relationship between 3D object and 2D image spaces. Accordingly, utilizing this formulation, and taking the occlusions into account, the information can be fused over all the images, therefore, increasing the accuracy of the verification process. It is iterative, because the location of 2D image features in the image is a nonlinear function of the position of their respective 3D model features and the viewpoint. Therefore, the solution is based on an iterative re-weighting least squares minimization. One can argue about this error criterion, as its use can be justified only when certain assumptions on the noise distribution of the measurements hold. However, in the case that this noise distribution is unknown using a least squares error criterion is a reasonable choice. Integrating a set of geometric and topologic constraints that force the estimation process to obtain a globally optimized solution refines the process. These constraints provide a large support for including only the relevant image information into the process. In addition, they allow FBMV to find photometrically weak or occluded image features that otherwise could not be fund without also finding many irrelevant features. It is a feature-based method because 3D positions of the model features are directly used as an initial guess to guide the search for finding homologous 2D image features during the estimation process in all corresponding images.

Each of the building hypotheses is evaluated using the estimation algorithm. The result of estimation can be used as a decision criterion in the verification phase. For example, by setting a threshold on the residuals, one can make a decision about the correctness of the model primitives. The correction is driven by the difference between the reconstructed coarse model and measured features within images. An essential point in such an

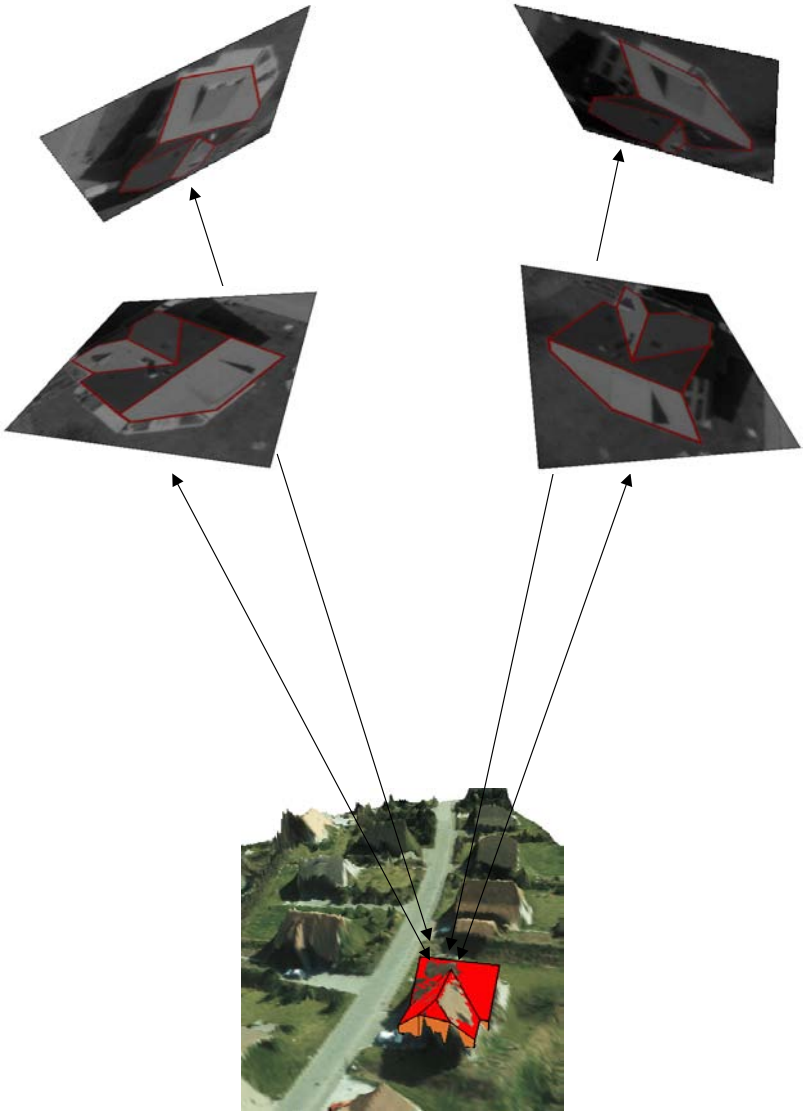


Figure 6.1: Perspective projection of a building into corresponding images

estimation algorithm is the comparison level used. It is only possible to calculate such a difference when the hypothesis model and measurement are at the same level of comparison. The choice of the comparison level is dependent on many factors (Schutte & Hilhorst 1993). The key issues that must be taken into account are the possible loss of information in data reduction, and the introduction of modeling errors in the hypothesis model. An obvious choice is the level of the original measurement. For computer vision applications this is the pixel level. Although the possible loss of information in pixel level is minimized a few disadvantages exist for using the raw pixel level.

- The intensity of a pixel depends not only on the geometry of the scene, but also on the radiance of the surfaces, and non-geometric camera and light source parameters. This means that a radiance part needs to be added to the estimation model, which is not strictly necessary, as we are interested in the geometry only.

- The data set size of the measurement is quite large in the application considered. The digitized aerial photographs used typically have a size of several mega pixels. A prediction on the same level will also involve the same data set size. In addition, a few iterations will be needed until convergence is reached. This results in quite expensive computational cost.

In FBMV method, the explicit geometry of the straight-line in object model is used as a comparison level against an implicit description of the lines in image space. In addition the topological and geometrical properties of the planar surfaces are used as supporting constraints. An example of such a topological information is the adjacency relationships that provide implicit information about connectivity of the adjacent edges. The levels of comparison in object space are reached by a bottom-up segmentation of the images, as described in chapter 4, while the corresponding level in image space are derived by a pre-analysis of the magnitude and orientation of the pixels gray value gradient which is discussed in next section.

Figure 6.1 illustrate schematically the basic concepts of FBMV method. The process starts with a hypothesis object model, to be specific, the b-rep of the 3D reconstructed polyhedral-like building model. All the 3D model edges are back projected into the corresponding images taken from different viewpoints to define the homologous 2D model edges in images. The projected 2D model edges serve as initial guess to guide the estimation process. A buffer of uncertainty as a search space is generated around each 2D edge. Within the generated buffer, probable edge-pixels are selected based on the analysis of the direction and magnitude of the pixel gray value gradient.

The estimation process is an orthogonal linear least squares regression problem with the objective to simultaneously minimize the perpendicular sum of the Euclidean distances between the selected pixels to the projected 2D model edges in all the images. Therefore, the treatment for the robust parameters estimation for outliers detection and self diagnosis, discussed in chapter 3 are also applicable here. To support the minimization process, a set of constraint is also integrated into the estimation model. The purpose, description and a mathematical formulation of all the observations used in FBMV is discussed in the following sections.

6.4 Mathematical Foundation

The objective of this section is to formulate the verification of the hypothesis building model. This is carried out by back projecting the 3D coarse model into the corresponding 2D images. Although this transformation is a non-linear operation it is a smooth and well-behaved transformation, and it is a promising candidate for the application of the well known Gauss-Markov estimation model based upon an iterative least squares minimization error criterion. This method requires the appropriate initial guess for the unknown parameters. These values are provided by the geometric and topological information derived from the reconstructed coarse model itself. Since the model primitives are projected, manipulated and the new values estimated in the inner loop of the matching process, it is important that possible and efficient sources of information particular to the estimation problem will be exploited. In practice, this is done by dividing the whole spectrum of the observations derived from the building model's description into three major categories as *image based, object based* and *image-object based observation equations* which are discussed in the next subsections. At the starting point of the estimation process a dense internal data structure is built from the model description. The structure is used to define identical 3D points, edges, and planar surfaces, as well as their topological relationships. In this manner, the model primitives may move independently while being attached to their adjacent primitives. In this way, an edge element connecting two model points can stretch under the influence of shifting one of its endpoint from its initial location and rotate under the influence of the movement of the another endpoint. In order to preserve consistency, the following notations are defined identically through the formalization of the method.

- X, Y, Z; the coordinates of the model points in 3D object space,
- x, y; the coordinates of the projected model points in 2D image space,
- $i \geq 3$; stands as the model points ID or representing the index of selected edge pixels in 2D images,
- $j \geq 3$; represents the model edges ID,
- $k \geq 1$; indicates the model face(s) ID,
- $r \geq 2$; represents different aerial images,

In addition, the following assumption should be taken into account:

- **Known parameters**; The precise exterior and interior orientation parameters of all the images are given. Therefore, extraction of 3D terrain coordinates is simply feasible through classical photogrammetric methods.

- **Approximate values**; The initial pose, number, and topological relationships such as connectivity and adjacency of the geometric primitives of the roof structure, i.e., 3D roof edges or 3D plane-roof polygons and their mathematical descriptions, such as surface normal, edge direction, etc., are known or can be computationally obtained from the reconstructed coarse building model. These initial values are updated in each iteration.

- **Unknown parameters**; The ultimate goal of the estimation model is to define the precise position of the 3D roof points which subsequently are used to define the geometrical description of the related geometric primitives of the fine polyhedral-like building model. However, during the estimation process the parameters of the 2D image edges are also considered as unknowns.

6.4.1 Image Based Observations

The observations concerned in this class are introduced into the estimation process for solving the unknown parameters of 2D primitives such as the parameters of the 2D image edges or the coordinates of the homologous 2D model points in image space. Two types of observations 1) *linearity* which serves as functional model of the estimation process, and 2) *connectivity* which is applied as topological constraint are integrated into the system as image based observations and are discussed next.

6.4.1.1 Linearity: A Local Internal Geometric Constraint

As we have mentioned, the functional model of the estimation process is a linear regression problem with the objective to minimize the orthogonal distances between the selected edge pixels and projected 2D model edges in image space, which is supported by additional constraints. In fact, it would be sufficient to simply solve a resection equation for modifying the coarse building model if we are able to find the corresponding matches between the model points and their homologous points in the respective images. To overcome the problem of feature correspondence, the match is actually established between the projected model edge and partial edges or alternatively edge-pixels in the image. In other words, since the precise position of the endpoints of image edge is unknown or is missing due to the occlusion, it is necessary to minimize only the perpendicular distance from representative points on an image edge to the projected model edge.

An edge model E_j of a coarse polyhedral-like building model in 3D object space is approximated by a straight line and represented in a parametric form as:

$$E_j : \quad P_i = P_j^{(o)} + s_j t_i \qquad ; \qquad t_i \in [t_{start}, t_{end}] \tag{6.1}$$

where $P_j^{(o)}$ represent a reference point on the line, P_i is an arbitrary point on the edge E_j, s_j represents direction of the line, t_i is a bounded real number corresponding to the point P_i, and t_{start}, t_{end} are minimum and maximum value of t_i defined by the start and end point of the edge E_j respectively. However, in order to measure the perpendicular distance from a representative edge-pixel (x_i, y_i) to the projected 2D model edge e_j, it is useful to express the projected edge in image space in the following form:

$$e_j : \qquad x_i \sin\theta - y_i \cos\theta - d = 0 \tag{6.2}$$

where d is the distance of the origin from the line, θ expresses the angle between the edge and x axis (or the angle between the edge normal and y axis). In practice, the initial parameters of the 2D edge (θ_j, d_j) are obtained by back projecting the two endpoints of the 3D model edge into the corresponding images using the collinearity equation (6.10). The computed 2D points are then plugged into the equations (6.3), (6.4), in order to derive the initial 2D edge parameters.

$$\theta_j \quad = \quad \arctan \frac{y_{end} - y_{start}}{x_{end} - x_{start}} \tag{6.3}$$

$$d_j \quad = \quad x_{start} \sin\theta_j - y_{start} \cos\theta_j \tag{6.4}$$

The projected 2D edge $e_{(r,j)}$ in image I_r serves as initial guess for finding the exact position of the edge model within the corresponding images. An uncertainty buffer with a user specified width is generated around each edge model in image space based on the initial position of the edge and is used as the search space to find the representative edge-pixels. Figure 6.2 shows the generated buffer around the homologous model edges of a reconstructed coarse building model in four images taken from different view points.

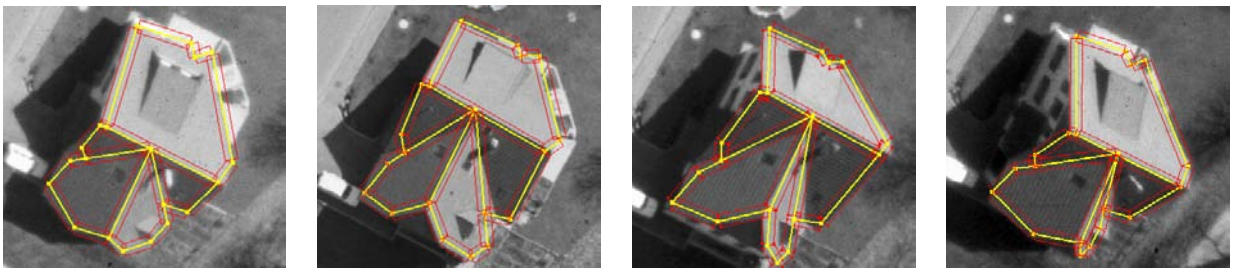


Figure 6.2: Uncertainty buffer of homologous model edges in corresponding images

Each pixel within the specified buffer is selected as a representative edge-pixel if it satisfies the following two conditions:

- its gradient direction will approximately be perpendicular to the edge model direction, and
- the magnitude of its gray value gradient will be more than a data-driven adaptive threshold. This threshold is computed based on a cumulative histogram of the gradient magnitude of all the candidate pixels within the buffer, which satisfy the first criteria. The gradient magnitude associated with each selected pixel is used as its weight in the estimation model.

Applying these two conditions for the selection of representative edge-pixels which should satisfy equation (6.2) has the advantages that firstly pixels which are laid on the edge image have stronger effect during the fitting procedure because they have more weight in the estimation process. Secondly, the gradients caused by background objects will not interfere with the parameter estimations, as they are not in the approximate direction of the edge model. Figure (6.3) indicates the selected edge-pixels of the homologous 2D model edges in corresponding images within the generated uncertainty buffer in the first iteration of the estimation process.

Figure 6.3: Selected edge-pixels during the first iteration of the estimation process

At this point, after selecting the edge-pixels representative, we are ready to introduce the linearity constraints into the estimation model. Let us once more represent the equation of the projected 2D edge in image I_r, passing through the selected edge-pixel (x_i^{img}, y_i^{img}) in the following form:

$$f_{(r,j)}(\theta, d) = x_i^{img} \sin\theta_{(r,j)} - y_i^{img} \cos\theta_{(r,j)} - d_{(r,j)} = e_i(x_i^{img}, y_i^{img}) \tag{6.5}$$

where $d_{(r,j)}$ is the distance of the origin from the edge, and $\theta_{(r,j)}$ expresses the edge angle respect to x axis. In this formulation, the orthogonal distance e_i represents an added error parameter, which acts as a cost function and should be minimized during the estimation (see figure 6.4).

Linearization of the equation (6.5) with respect to its parameters $(d_{(r,j)}, \theta_{(r,j)})$ results in the following formulation:

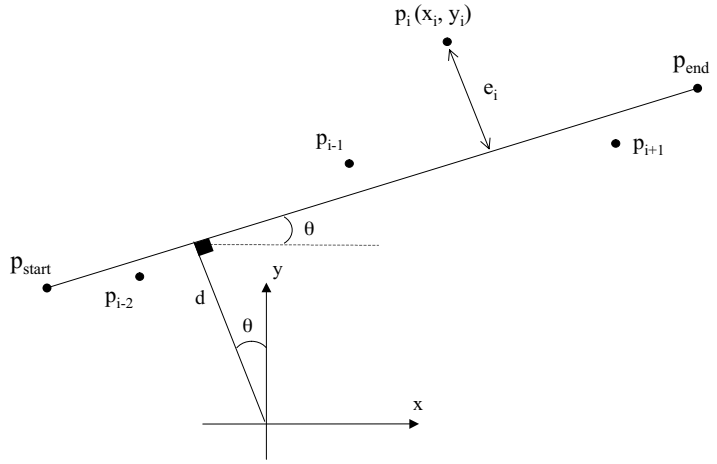


Figure 6.4: Regression of a 2D image edge to the representative edge-pixels

$$\frac{\partial f_{(r,j)}}{\partial \theta_{|\theta=\theta^0_{(r,j)}}}\Delta\theta_{(r,j)} + \frac{\partial f_{(r,j)}}{\partial d_{|d=d^0_{(r,j)}}}\Delta d_{(r,j)} - l_i = e_i(x_i^{img}, y_i^{img}) \tag{6.6}$$

where

$$l_i \quad = \quad d^0_{(r,j)} - x_i^{img(0)} \sin\theta^0_{(r,j)} + y_i^{img(0)} \cos\theta^0_{(r,j)}.$$

For every selected edge-pixel (x_i^{img}, y_i^{img}) of each 2D edge model $e_{(r,j)}$, within every image I_r, an equation of type (6.6) is inserted into the system of equations. The total system of equations can be written in matrix form as:

$$\mathbf{A}_{linear} \cdot \mathbf{x} - \mathbf{l}_{linear} = \mathbf{e} \quad ; \quad \sigma_0^2 \mathbf{P}^{-1}_{linear}. \tag{6.7}$$

The $\mathbf{l}_{linear}$, is the observation vector containing the orthogonal distance between the candidate pixels and their respective 2D model edge in image space. $\mathbf{x}$ is the vector of unknowns consisting of the correction of the edge parameters $(\Delta\theta_{(r,j)}, \Delta d_{(r,j)})$, $\mathbf{A}_{linear}$ is the associated design matrix including derivatives of the observation equations with respect to the unknowns. The matrix $\mathbf{P}_{linear}$, is the corresponding weight matrix which is introduced as a diagonal matrix and is determined based on the normalized gradient magnitude of each candidate pixel, and $\mathbf{e}$ is a error vector with the statistical assumption:

$$E(\mathbf{e}) = 0.$$

The system of (6.7) is the well known Gauss-Markov estimation model. The least squares estimation in this model gives a unique and most probable set of estimates for all the parameters of the 2D model edges.

To make the selection process robust and impose a self-diagnosis mechanism, the generated buffer is updated in a regular interval during the iteration process. As the process is iterated, the initial parameters of the 2D edges are updated based on the minimization of the orthogonal distance error between the selected edge-pixels and their respective 2D edges. Consequently the updated parameters define a new orientation for the generated buffer. In addition by introducing a smaller width, the size of the buffer is reduced. As a consequence, as outliers are excluded from the estimation process, this procedure reduces the computational burden and increases the accuracy and speed up the convergence process. Figure (6.5) indicates the selected edge-pixels of the corresponding 2D edges in the figure (6.3) for the last iteration of the estimation process.

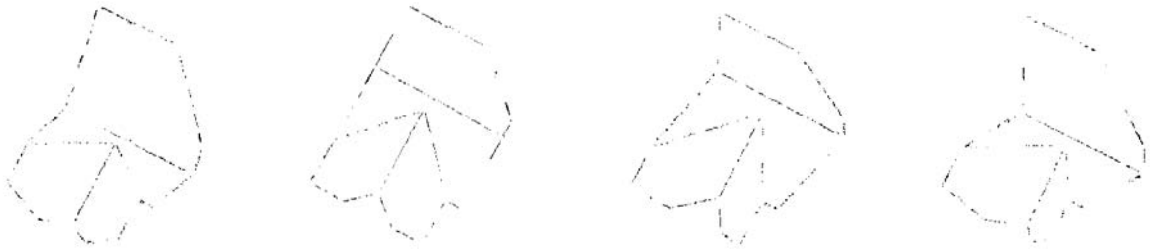


Figure 6.5: Selected edge-pixels during the last iteration of the estimation process

6.4.1.2 Connectivity: A Global Internal Topological Constraint

The connectivity constraints are integrated into the estimation model as a topological constraint based on the intersection point between adjacent model edges. Figure (6.6) depicts a corner of the building model when two edges $edge_1$ and $edge_2$ are connected through the intersection point p_{int}.

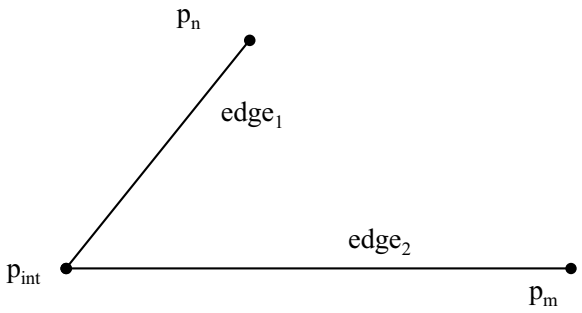


Figure 6.6: Intersection of two adjacent edges

In general, it is possible to introduce the connectivity constraints into the estimation process both in object space or image space, because it is invariant under the geometric transformation and it is independent of the embedded space. In this model, it is categorized as an image-based observation, in order to overcome the problem of correspondence between the model points within the images. In other words, by setting up the following formulation, the location of the corresponding 2D model points in different images is introduced implicitly based on the topological information, not the geometrical one. That means we do not compute the location of the intersection point explicitly based on the intersection of the two adjacent edges. Therefore, the problem of finding homologous points in respective images is not encountered, as it is required in the feature-based matching techniques. In fact, if we had the correspondence relationships between the homologous model points in different images, then the verification of the coarse model would be done simply by obtaining the exact location of the 3D model points based on the simple resection technique such as MPGC (Baltsavias 1991, Gruen & Stallmann 1991).

To formulate the connectivity constraints between adjacent edges for every associated edge member of a model point the following observation is introduced into the total system of equations. Once again, let us consider the equation (6.5), to represent a 2D edge $e_{(r,j)}$, in image I_r. Linearization of this equation with respect to its parameters, in this case 2D edge parameters $(d_{(r,j)}, \theta_{(r,j)})$ and 2D coordinates $(x_{int}^{img}, y_{int}^{img})$ of the intersection point in image space results in the following formulation:

$$\frac{\partial f_{(r,j)}}{\partial \theta_{|\theta=\theta^0_{(r,j)}}} \Delta\theta_{(r,j)} + \frac{\partial f_{(r,j)}}{\partial d_{|d=d^0_{(r,j)}}} \Delta d_{(r,j)} + \frac{\partial f_{(r,j)}}{\partial x_{|x=x_{int}^{img(0)}}} \Delta x_{int}^{img} + \frac{\partial f_{(r,j)}}{\partial y_{|y=y_{int}^{img(0)}}} \Delta y_{int}^{img} - l_{int} = e_i(x_{int}^{img}, y_{int}^{img}) \tag{6.8}$$

where

$$l_{int} = d^0_{r,j} - x^{img(0)}_{int} \sin\theta^0_{(r,j)} + y^{img(0)}_{int} \cos\theta^0_{(r,j)}.$$

The arrangement of the above equation in the form of a Gauss-Markov model for all the connected edges in corresponding images is expressed as follow:

$$\mathbf{A}_{connect} \cdot \mathbf{x} - \mathbf{l}_{connect} = \mathbf{e} \quad ; \quad \mathbf{P}_{connect} \tag{6.9}$$

where $\mathbf{l}_{connect}$ is the observation vector, $\mathbf{x}$ is the vector of unknowns consisting of the corrections of the 2D edge parameters $(d_{(r,j)}, \theta_{(r,j)})$ and corrections of the coordinates of the intersection point $(\triangle x^{img}_{int}, \triangle y^{img}_{int})$, $\mathbf{A}_{connect}$ is the associated design matrix consisting of the partial derivatives of the edge equation with respect to its unknown parameters, $\mathbf{P}_{connect}$ is the corresponding weight matrix, and $\mathbf{e}$ is the added error vector.

6.4.2 Image-Object Based Observations

These types of observations are the essential parts of the estimation model. They are integrated into the estimation process in order to establish the required link between the image and object space. They are acting as a bridge to tie the estimated corrections of the unknown parameters obtained in image space to their respective model parameters in object space during the iteration.

6.4.2.1 Collinearity: A Global External Geometric Constraint

The mapping relation between a point in 3D object space $P_i(X, Y, Z)$ and its perspective projection in 2D image space $p^{img}_i(x, y)$ can be represented by the classical collinearity equations as follows:

$$x^{cam}_i + F^x_i(X, Y, Z) = 0$$

$$y^{cam}_i + F^y_i(X, Y, Z) = 0$$

where

$$F^x_i(X, Y, Z) = f\frac{a_{11}(X_i - X_o) + a_{12}(Y_i - Y_o) + a_{13}(Z_i - Z_o)}{a_{31}(X_i - X_o) + a_{32}(Y_i - Y_o) + a_{33}(Z_i - Z_o)}$$

$$F^y_i(X, Y, Z) = f\frac{a_{21}(X_i - X_o) + a_{22}(Y_i - Y_o) + a_{23}(Z_i - Z_o)}{a_{31}(X_i - X_o) + a_{32}(Y_i - Y_o) + a_{33}(Z_i - Z_o)} \tag{6.10}$$

the $a_{11}, a_{12}, ..., a_{33}$ are the elements of the rotation matrix, X_o, Y_o, and Z_o are the location of the perspective center in object space, f is the principal distance of the sensor, and (x^{cam}_i, y^{cam}_i) are the coordinates of the point in camera coordinate system. Additionally, a mapping relation is needed to relate a point in camera system $p^{cam}_i(x_i, y_i)$ to its corresponding point in image coordinates system $p^{img}_i(x_i, y_i)$. The transformation parameters are expressed in the terms of an affine transformation:

$$x^{cam}_i = c_{11} \cdot x^{img}_i + c_{12} \cdot y^{img}_i + c_{10}$$

$$y^{cam}_i = c_{21} \cdot x^{img}_i + c_{22} \cdot y^{img}_i + c_{20} \tag{6.11}$$

where $c_{11}, c_{12}, c_{21}, c_{22}$ and c_{10}, c_{20} are the rotation and translation parameters of the affine transformation respectively. Assuming the interior and exterior orientation parameters of each image are given, then the unknowns

to be determined are the image coordinates (x_i^{img}, y_i^{img}) of the model point and its corresponding coordinates (X_i, Y_i, Z_i) in the 3D object space. Therefore, if the coordinates of a point is given in object space, the corresponding image coordinates of the point is simply derived from the equations (6.10) and (6.11), or alternatively, if an object is imaged from more than one viewpoint and the interior and exterior orientation parameters of the images are given, then 3D coordinates of the point in object space can be reconstructed by simultaneous intersection of the above collinearity conditions (resection in space). The collinearity equations imply that the location of 2D image features is a nonlinear function of the position of their respective 3D model features and viewpoint. The linearization of this equation with respect to its unknown parameters result in a equation that is a linear combination of the partial derivatives of the location of a point in camera space and its corresponding position in 3D object space, and is defined by:

$$\Delta x_i^{cam} + \frac{\partial F_i^x}{\partial X_{|X=X_i^0}} \Delta X_i + \frac{\partial F_i^x}{\partial Y_{|Y=Y_i^0}} \Delta Y_i + \frac{\partial F_i^x}{\partial Z_{|Z=Z_i^0}} \Delta Z_i + x_i^{cam(0)} + F_i^x(X^0, Y^0, Z^0) = 0$$

$$\Delta y_i^{cam} + \frac{\partial F_i^y}{\partial X_{|X=X_i^0}} \Delta X_i + \frac{\partial F_i^y}{\partial Y_{|Y=Y_i^0}} \Delta Y_i + \frac{\partial F_i^y}{\partial Z_{|Z=Z_i^0}} \Delta Z_i + y_i^{cam(0)} + F_i^y(X^0, Y^0, Z^0) = 0 \qquad (6.12)$$

the partial derivatives of equation (6.11) result in:

$$\begin{aligned} \Delta x_i^{cam} &= c_{11} \Delta x_i^{img} + c_{12} \Delta y_i^{img} \\ \Delta y_i^{cam} &= c_{21} \Delta x_i^{img} + c_{22} \Delta y_i^{img}. \end{aligned} \qquad (6.13)$$

Hence, the linearized equations concerning a 2D point in image coordinates $p_i^{img}(x, y)$ with respect to its 3D position $P_i(X, Y, Z)$ can be formulated by plugging the equations (6.13) into the equation (6.12) as:

$$c_{11} \Delta x_i^{img} + c_{12} \Delta y_i^{img} + \frac{\partial F_i^x}{\partial X_{|X=X_i^0}} \Delta X_i + \frac{\partial F_i^x}{\partial Y_{|Y=Y_i^0}} \Delta Y_i + \frac{\partial F_i^x}{\partial Z_{|Z=Z_i^0}} \Delta Z_i - l_i^x = e_i^x$$

$$c_{21} \Delta x_i^{img} + c_{22} \Delta y_i^{img} + \frac{\partial F_i^y}{\partial X_{|X=X_i^0}} \Delta X_i + \frac{\partial F_i^y}{\partial Y_{|Y=Y_i^0}} \Delta Y_i + \frac{\partial F_i^y}{\partial Z_{|Z=Z_i^0}} \Delta Z_i - l_i^y = e_i^y \qquad (6.14)$$

where

$$\begin{aligned} l_i^x &= -(x_i^{cam(0)} + F_i^x(X^0, Y^0, Z^0)) \\ l_i^y &= -(y_i^{cam(0)} + F_i^y(X^0, Y^0, Z^0)). \end{aligned}$$

The equation (6.14) for all the model points can be arranged into the matrix form in a Gauss-Markov model as:

$$\mathbf{A}_{collinear} \cdot \mathbf{x} - \mathbf{l}_{collinear} = \mathbf{e} \quad ; \quad \mathbf{P}_{collinear} \qquad (6.15)$$

where $\mathbf{l}_{collinear}$, is the observation vector containing the difference between the coordinates of the initial 2D model points computed by the collinearity equations and the one which is implicitly introduced to the total estimation model by connectivity equations. $\mathbf{x}$ is the vector of unknowns consisting the correction of the model points in object space $(\Delta X_i, \Delta Y_i, \Delta Z_i)$, and image space $(\Delta x_i^{img}, \Delta y_i^{img})$, $\mathbf{A}_{collinear}$ is the associated design matrix including derivatives of the observation equations with respect to the unknowns, $\mathbf{P}_{collinear}$ is the corresponding weight matrix, and $\mathbf{e}$ is an added error vector.

6.4.3 Object Based Observations

As it has been discussed so far, the main objective of the FBMV method is to integrate the model-driven information into the estimation model as supporting constraints. The observation equations that are classified in this category are directly obtained based upon the description of the reconstructed coarse building model, before or during the estimation process. These constraints are introduced between the model primitives in object space as global or local geometric constraints and are linearized and applied as weighted observation equations. In this manner, the integration of the model primitives as unknowns into the total system of equations is completely flexible. Introducing the relationship between the model primitives as a strict condition by increasing its weight or alternatively reducing its influence into the system by decreasing its weight.

6.4.3.1 Coplanarity: A Global External Geometric Constraint

Due to the pragmatic assumption that building roof structures are geometrically described by the aggregation of $k \geq 1$ planar surface(s), all the bounding points $P_i(X, Y, Z)$ of the 3D plane-roof polygon F_k, should satisfy the coplanarity condition defined as follows:

$$f_{(i,k)}(\vec{n}, D) = A_k X_i + B_k Y_i + C_k Z_i + D_k = e_i \tag{6.16}$$

where (A_k, B_k, C_k), are the components of the surface normal $\vec{n}_k$, D_k is the distance from origin to the plane-roof polygon F_k, and e_i is an added error parameter. The partial derivatives of this equation with respect to the unknown parameters, that is the 3D coordinates of the model points, is obtained by:

$$\frac{\partial f_{(i,k)}}{\partial X_{|X=X_i^0}} \Delta X_i + \frac{\partial f_{(i,k)}}{\partial Y_{|Y=Y_i^0}} \Delta Y_i + \frac{\partial f_{(i,k)}}{\partial Z_{|Z=Z_i^0}} \Delta Z_i - l_i = e_i \tag{6.17}$$

where

$$l_i = -(A_k^0 X_i^0 + B_k^0 Y_i^0 + C_k^0 Z_i^0 + D_k^0).$$

In fact, the corrections $(\Delta X_i, \Delta Y_i, \Delta Z_i)$ in each iteration represent changes to the initial location of the model points, while the best planar fit to the updated model points is obtained. Introducing an equation of the type (6.17) for every point of the plane-roof polygons F_k in the estimation model and arranging all the equations in the matrix form result a Gauss-Markov model as:

$$\mathbf{A}_{coplanar} \cdot \mathbf{x} - \mathbf{l}_{coplanar} = \mathbf{e} \quad ; \quad \mathbf{P}_{coplanar} \tag{6.18}$$

where $\mathbf{l}_{coplanar}$ is the observation vector, $\mathbf{x}$ is the vector of unknowns consisting the corrections of the coordinates of the model points $(\Delta X_i, \Delta Y_i, \Delta Z_i)$, $\mathbf{A}_{coplanar}$ is the associated design matrix determined by the initial values of the surface normal $\vec{n}_k^0$, and $\mathbf{P}_{coplanar}$ is the corresponding weight matrix.

It should be mentioned that after each iteration the equation (6.16) is solved with the improved model points coordinates in order to obtain a new estimate for the surface normal $\vec{n}_k$, and the scalar value D_k. The estimated parameters are used in the next iteration as the new initial values.

6.4.3.2 Conditional Constraints

The FBMV is an iterative procedure based on Newton-Raphson method, thus it converges to the minimum and becomes closer to the correct solution in every iteration, unless the system is degenerated, the initial values are so far away from the true solution, or the estimation model is incorrectly established. This property enables us to integrate additional constraints between the model primitives during the iteration process, if certain conditions are satisfied. As we have mentioned previously the strength of our method is that it works based on a data-driven generic data model. That means instead of imposing certain regularities or conditions into the

model in the earlier stages of the reconstruction process, these regularities and constraints are introduced into the model in the higher level process of reconstruction. Such constraints are the *orthogonality*, or *parallelity* between the adjacent model edges, *symmetricalness* or *semi-symmetricalness* between the adjacent faces, and so on. The decision to impose these constraints into the estimation model is made during model verification process when the required criteria are met. The triggered constraints are integrated into the model, simply by adding a new row to the total system of equations. For the sake of completeness the orthogonality constraints are elaborated in details next, the other constraints can be dealt with in the same manner.

Orthogonality: A Local External Geometric Constraint
Figure (6.7) represent the angle α, between two 3D model edges E_1, and E_2. The conditional geometric constraint of the orthogonality is applied for every model point P_i, when α satisfies the following condition during each iteration:

$$90 - t \leq \alpha \leq 90 + t \tag{6.19}$$

where t is a threshold (e.g., 5°) indicating the small deviation of α from its expected value i.e. 90°. In fact, when two adjacent edges are considered orthogonal then the following constraints should be met:

$$f_i(X, Y, Z) = a_1 a_2 + b_1 b_2 + c_1 c_2 = 0 = e_i \tag{6.20}$$

where $\vec{E_1}(a_1, b_1, c_1)$, and $\vec{E_2}(a_2, b_2, c_2)$, are the directions of the E_1 and E_2 respectively and e_i is an added noise parameter.

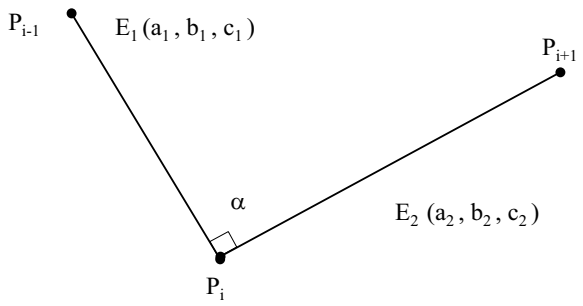


Figure 6.7: Two orthogonal adjacent edges

Linearization of the equation (6.20) with respect to the position of the model points $P_i(X, Y, Z)$ result in the form:

$$\frac{\partial f_i}{\partial X_{|X=X_i^0}} \Delta X_i + \frac{\partial f_i}{\partial Y_{|Y=Y_i^0}} \Delta Y_i + \frac{\partial f_i}{\partial Z_{|Z=Z_i^0}} \Delta Z_i - l_i = e_i \tag{6.21}$$

where

$$l_i = -f_i^0(X, Y, Z) = -(a_1^0 a_2^0 + b_1^0 b_2^0 + c_1^0 c_2^0).$$

Therefore, introducing an equation of the type (6.21) as a weighted observation equation for every model point and once again formulating them in a matrix form will result in the well known formulation of a Gauss-Markov model such as :

$$\mathbf{A}_{ortho} \cdot \mathbf{x} - \mathbf{l}_{ortho} = \mathbf{e} \quad ; \quad \mathbf{P}_{ortho} \tag{6.22}$$

where $\mathbf{l}_{ortho}$ is the observation vector consisting of the magnitude of the scalar product of the adjacent edges, $\mathbf{x}$ is the vector of unknowns consisting of the corrections to the coordinates of the model points $(\Delta X_i, \Delta Y_i, \Delta Z_i)$,

$\mathbf{A}_{ortho}$ is the associated design matrix determined by the partial derivatives of the equation (6.20) with respect to its unknown parameters, $\mathbf{P}_{ortho}$ is the corresponding weight matrix and $\mathbf{e}$ is the added error vector.

The other conditional constraints can be implemented in the same manner as discussed for the orthogonality constraint and integrated into the estimation model.

It should be mentioned that in practice, the *parallelity constraint* between the adjacent model edges is implemented in slightly different approach as we have discussed so far. Owing to the fact that in a b-rep of the object model (see chapter 5), we are concern to store and represent the model structure with essential geometric primitives and in a compact form, thus instead of forcing strictly two adjacent edges to become parallel, when they satisfy the parallelity constraint ($\alpha \approx 180°$), and both are only associated with one model face, the FBMV will merge them into one model edge primitive. In a similar way, during iteration process, a redundant edge model is removed if it coincides with one of its adjacent edges ($\alpha \approx 0°$). This is due to the fact that the specified edge is not representing an essential or real part of the object model and it has been considered part of the reconstructed coarse model because of the inefficiency of the segmentation process to detect only the relevant edge primitives.

6.4.4 Combined Least Squares Adjustment

Based on the assumption that all the parameters involved in the estimation model are considered observations, and consequently equations arising with the constraints or conditions are introduced into the total system of equations as weighted observation equations, then we are able to join the different estimation models which are formed by the (6.6), (6.8), (6.14), (6.17), (6.21), or any other equations into a unified combined Gauss-Markov model such as:

$$\sum \mathbf{A}_c\mathbf{x} - \sum \mathbf{l}_c = \sum \mathbf{e}_c \quad ; \quad \sum \mathbf{P}_c \tag{6.23}$$

where $\mathbf{x}$ represents the total vector of unknowns. The combined least squares solution of the equation (6.23) gives the vector of estimates for correction to the initial parameters of the model as follows:

$$\hat{\mathbf{x}} = (\sum \mathbf{A^T}_c\mathbf{P}_c\mathbf{A}_c)^{-1}(\sum \mathbf{A^T}_c\mathbf{P}_c\mathbf{l}_c). \tag{6.24}$$

Consequently, the vector of total residuals $\hat{\mathbf{e}}$, and a posteriori estimation of the variance factor $\hat{\sigma}_0^2$ can be computed by:

$$\hat{\mathbf{e}} = (\sum \mathbf{A^T}_c)\hat{\mathbf{x}} - (\sum \mathbf{l}_c) \tag{6.25}$$

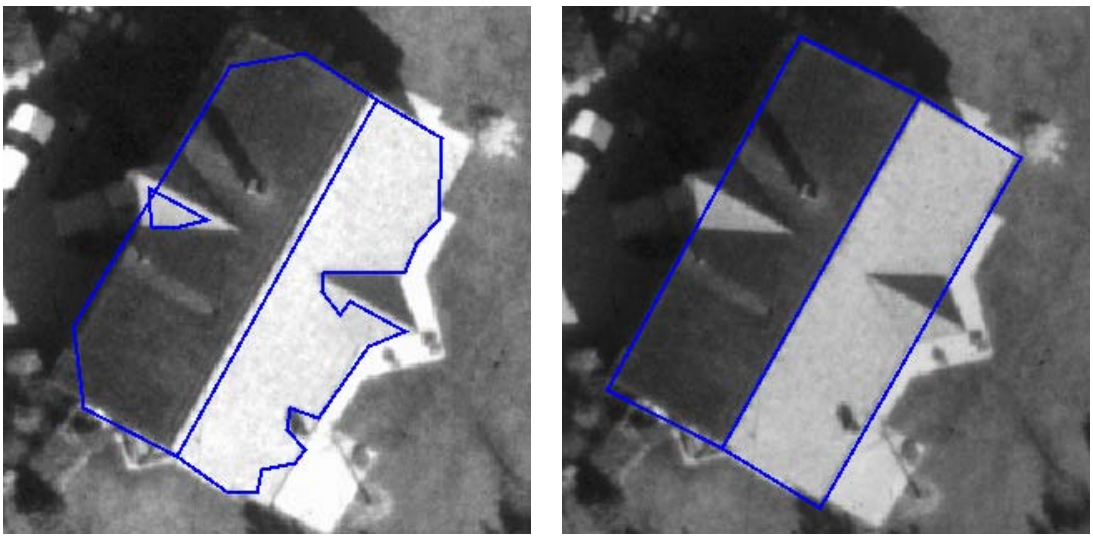
$$\hat{\sigma}_0^2 = \frac{\hat{\mathbf{e}}^\mathbf{T}(\sum \mathbf{P}_c)\hat{\mathbf{e}}}{n-u}. \tag{6.26}$$

Furthermore, the estimated reference variance $\hat{\sigma}_0^2$, may be compared to a priori value σ_0^2 using a chi-square χ^2 test, in order to assess the performance of the estimation (Mikhail 1976).

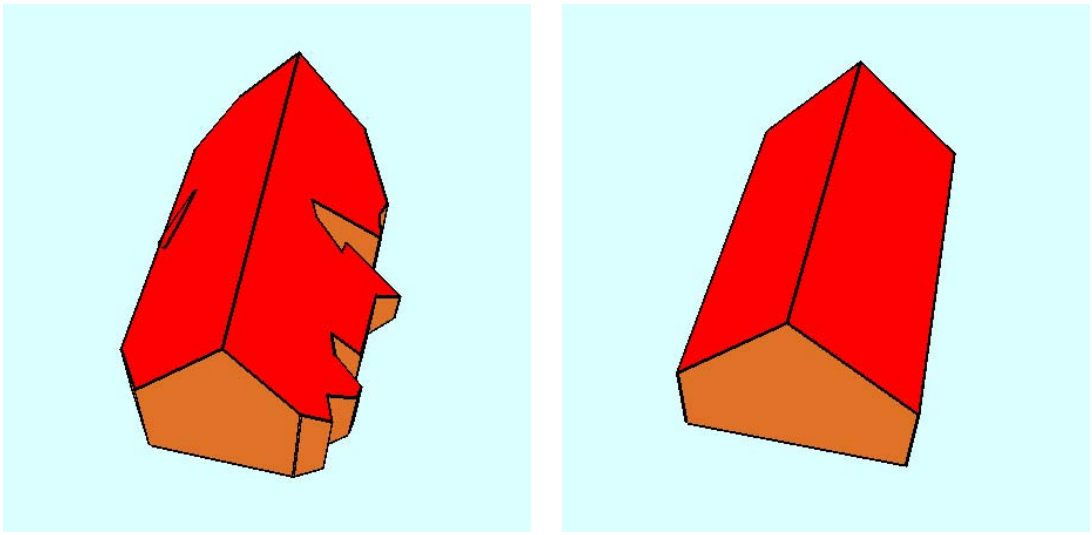
6.5 Experiments and Result

To evaluate the performance of the proposed method and to visualize the outcome of the FBMV algorithm, the three representative buildings of the Avenches data set are selected. It should be noted once more, that the whole process of model verification is only applied on the roof structure and the vertical walls are added to the building model at the final stage. This is performed based on the analysis of the bounding edges and points of the verified fine roof structures. The first building shown in figure (6.8) is a simple gable roof structure. The reconstructed coarse model (figure 6.8-c) indicates that the reconstruction process performed by POLYMODELER (see chap.5) recovered the fundamental structure of the building. However, due to the presence of a

dormer window on top of the roof and also the low contrast of the roof outline, the bounding edges of the building are broken into the small pieces and are very rugged. However, the FBMV approach successfully verified the model and removed the redundant edge segments. Furthermore, imposing the orthogonality constraints during the verification process enables FBMV to accurately recover the building corners (see figure 6.8-d), which were trimmed off during the segmentation process. Figures (6.8-a and 6.8-b) show the initial and modified building model overlaid on the corresponding aerial image, respectively.



a b



c d

Figure 6.8: Reconstructed gable roof structure building: a) reconstructed coarse building model overlaid on aerial image, b) reconstructed fine building model overlaid on aerial image, c) perspective view of the reconstructed 3D coarse building model, d) perspective view of the 3D fine building model.

The second example deals with reconstruction of a hipped-gable roof structure (see figure 6.9). As it is shown in figures (6.9-a and 6.9-c), although the major structures of the building, the bounding edges and even small protruding structure of the roof is reconstructed correctly. Due to the low contrast of the edge segments of the hip tile of the roof, the reconstructed edge is completely shifted away from its correct position. In addition, the position of the corner points is not precisely defined. The modification of the model based on a multi-photo estimation process and imposing the global constraints as discussed previously, enables FBMV to recover this occluded part of the roof and forces the displaced model edge to located in its true position. Furthermore it also define the positions of the model points more accurately (see figures 6.8-b and 6.8-d).

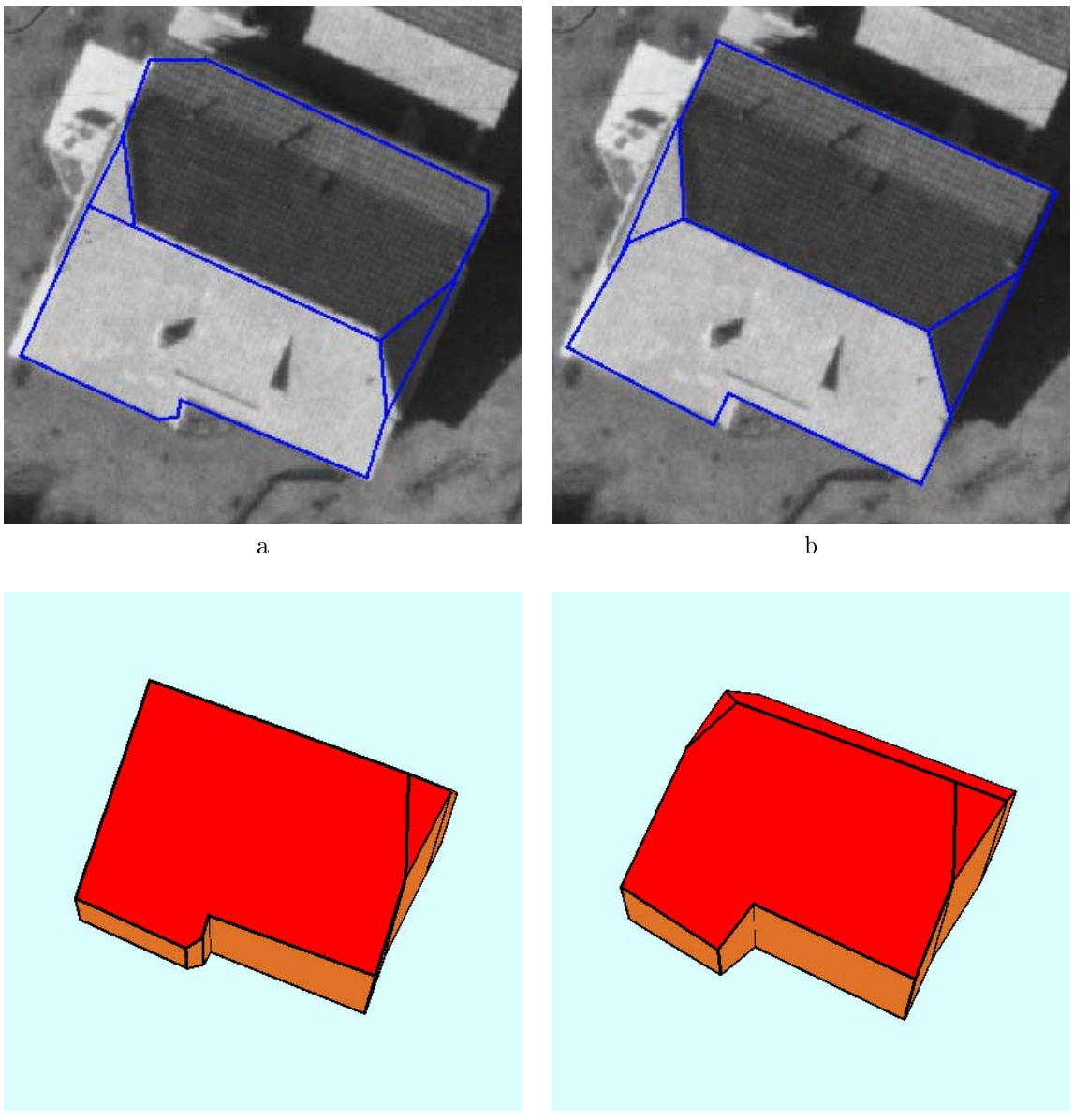
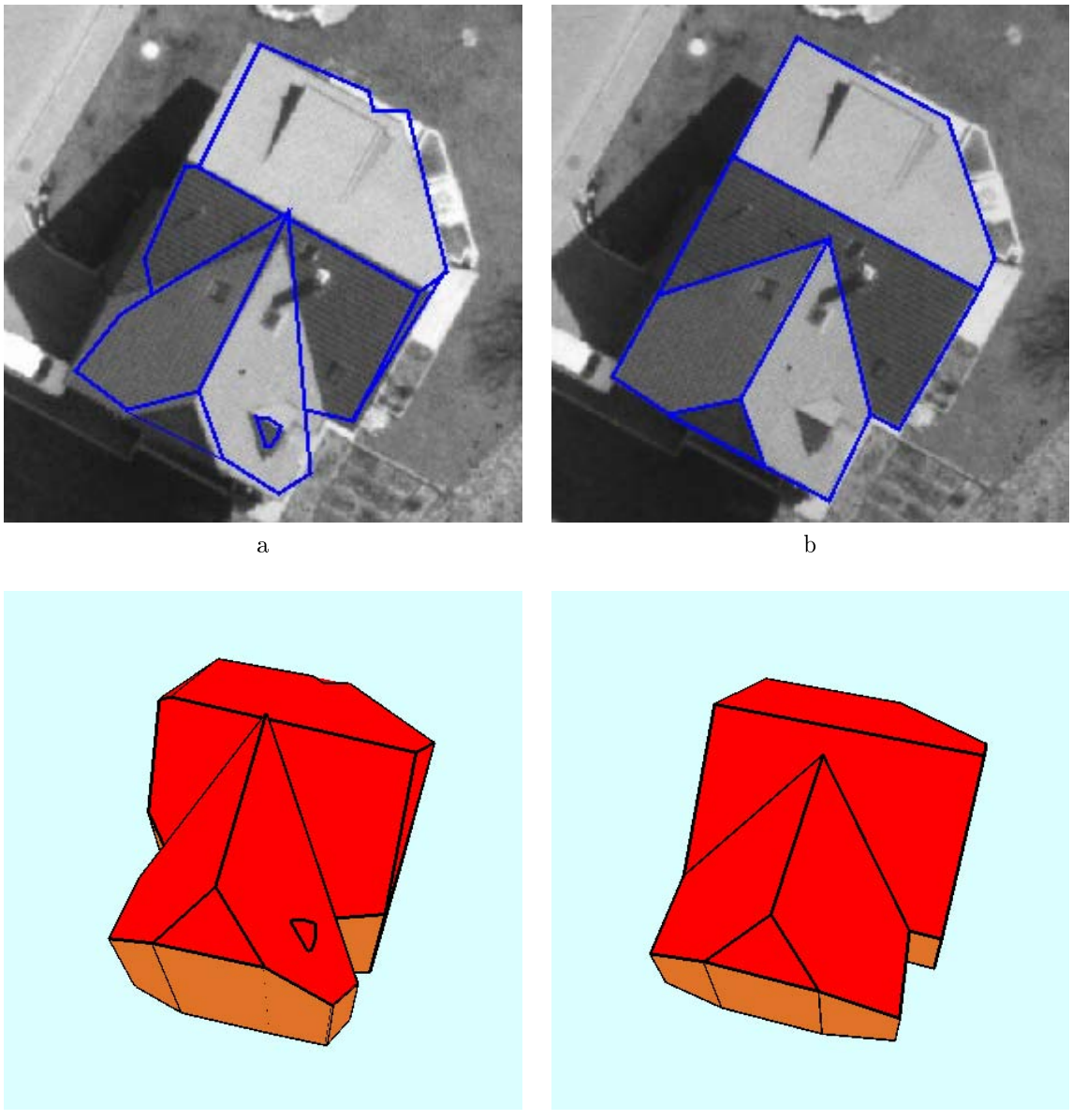


Figure 6.9: Reconstructed hipped-gable roof structure building: a) reconstructed coarse building model overlaid on aerial image, b) reconstructed fine building model overlaid on aerial image, c) perspective view of the reconstructed 3D coarse building model, d) perspective view of the 3D fine building model.

The last example shown in figure (6.10) represents a more complex roof structure. Although the hypothesis coarse building model generated by POLY-MODELER describes the building completely. Still there are displacements in some of the model primitives, specially the intersection point between three adjacent plane-roof polygons is shifted significantly from its real position (see figure 6.10-a). This is due to the failure in defining the correct orientation (slope) of the respective 3D plane-roof polygons in space, which is caused by the low quality of the utilized DSM. By applying the FBMV process, the normal vectors $\vec{n}_k$ of every plane-roof polygons are recovered precisely and consequently the intersection point is moved to its real position. In addition, the real bounding edges of the model are also precisely located and the redundant one is eliminated from the final model (figure 6.10-d).



a b

c d

Figure 6.10: Reconstructed complex roof structure building: a) reconstructed coarse building model overlaid on aerial image, b) reconstructed fine building model overlaid on aerial image, c) perspective view of the reconstructed 3D coarse building model, d) perspective view of the 3D fine building model.

The above experimental results show the strength and generality of the proposed FBMV, in recovering the reliable and accurately defined geometric primitives of different sorts of the building structures, which is an essential part of any automated vision system. It shows that the proposed method is capable of working with any complex polyhedral-like object model, if an appropriate initial hypothesis model is available. To complete

the performance evaluation of the FBMV, the following section is dedicated to numerical analysis and assessment of the quality of the final model obtained from the estimation model.

6.6 Quality Assessment

One of the key issues of the FBMV method is its ability to provide the essential tools for evaluation of the quality of the reconstructed model and its geometric primitives. As discussed previously, the least squares solution provides an estimate for the variance factor $\hat{\sigma}_0^2$ which can be used for the performance evaluation of the estimation process. In other words, it is used to judge whether or not the estimation model is consistent with the earlier assumption that the noise distribution follows a normal distribution function with a given standard deviation, which was the motivation to apply a least squares minimization of the error criterion. In addition, considering sufficient agreement between the estimation model and our early assumption, the standard and statistically well known covariance matrix $\hat{D}(\hat{\mathbf{x}})$ of the estimated parameters can be obtained as follows:

$$\hat{D}(\hat{\mathbf{x}}) = \hat{\sigma}_0^2 (\sum \mathbf{A}^{\mathbf{T}}{}_c \mathbf{P}_c \mathbf{A}_c)^{-1} \tag{6.27}$$

The estimated variances of the unknown parameters, specifically in our case the coordinates of the model points in 3D space $(\hat{\sigma}_X^2, \hat{\sigma}_Y^2, \hat{\sigma}_Z^2)$ are the qualitative measures which indicate the accuracy of the model primitives and act as the decision criteria in order to reject or accept the estimated model elements based on the simple thresholding process. The evaluation process can be integrated into the whole chain of reconstruction process as an edition process (traffic light concept (Förstner 1996)). In a simple manner, these measures give a hint to the end user to perform a visual check on the end product and perform the required modifications on the signalized model primitives if necessary. In the following are the numerical results and the statistical analysis of a single building shown in figure (6.11).

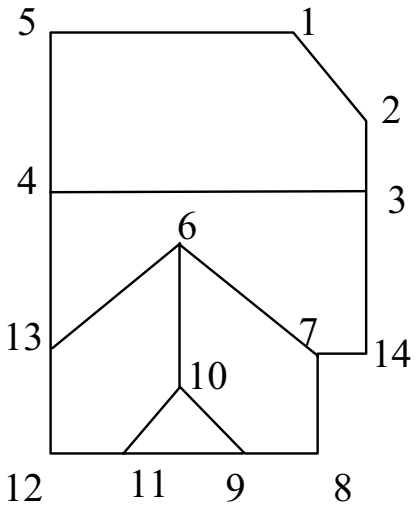


Figure 6.11: Top view of a single complex building

The test was carried out for the verification of the model based on utilizing two and four corresponding images taken from different views. In addition, the reconstructed building in every test is compared with a reference model digitized manually by an operator, in order to show a realistic quality measure of the process as well. The results are tabulated in tables (6.1), and (6.2) respectively.

A comparison of the estimated variances of the model points coordinates with respect to the absolute values of the differences between the estimated coordinates and the reference coordinates deduce consistency and agreement in both tests. Moving from the estimation model based on two images toward the one utilizing four images indicates a tendency in increasing the accuracy of the estimated model points, as it was expected. In addition, imposing more images into the estimation process increase the reliability of the model.

Point-ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	RMSE
ΔX	0.26	0.20	0.09	0.17	0.20	0.10	0.17	0.41	0.16	0.16	0.23	0.12	0.38	0.18	0.22
ΔY	-0.02	-0.11	-0.04	-0.16	-0.02	-0.09	-0.28	0.14	-0.15	-0.08	0.01	0.07	-0.21	0.30	0.15
ΔZ	0.03	0.03	0.12	-0.17	0.05	0.10	-0.18	0.22	0.03	0.28	-0.03	-0.26	-0.35	-0.17	0.18
σ_x^2	0.44	0.44	0.65	0.65	0.30	0.21	0.61	0.07	0.58	0.23	0.37	0.37	0.37	0.37	
σ_y^2	0.71	0.70	0.23	0.23	0.35	0.35	0.37	0.54	0.54	0.64	0.42	0.43	0.45	0.11	
σ_z^2	0.69	0.68	0.16	0.16	0.26	0.12	0.33	0.20	0.62	0.15	0.30	0.31	0.24	0.08	

Table 6.1: Verified coarse building model based on two corresponding aerial images

Point-ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	RMSE
ΔX	0.19	0.20	0.19	0.17	0.20	0.17	0.11	0.28	0.03	0.10	0.23	0.12	0.38	0.06	0.19
ΔY	-0.09	-0.16	-0.12	-0.20	-0.10	-0.17	-0.27	-0.17	-0.18	-0.19	-0.02	-0.01	-0.18	-0.29	0.17
ΔZ	-0.15	-0.15	-0.06	-0.23	-0.10	0.08	-0.13	0.09	-0.14	0.04	-0.27	-0.23	-0.19	-0.13	0.16
σ_x^2	0.34	0.34	0.49	0.49	0.20	0.23	0.64	0.08	0.56	0.23	0.22	0.22	0.30	0.26	
σ_y^2	0.30	0.30	0.23	0.23	0.23	0.32	0.36	0.30	0.32	0.46	0.32	0.32	0.39	0.11	
σ_z^2	0.34	0.34	0.21	0.21	0.26	0.16	0.46	0.17	0.64	0.20	0.25	0.25	0.26	0.09	

Table 6.2: Verified coarse building model based on four corresponding aerial images

It is discussed in the previous sections that the estimation process can improve significantly, if the robust M-estimator proposed in chapter 3, is integrated into the process of FBMV. In this manner the remaining outliers, undetected during selection of candidate pixels based on uncertainty buffer can be further filtered out from the process.

It should be stressed that most of the imposed constraints and their internal relationships can be altered and introduced with different formulations. In addition, more sophisticated robust techniques for outlier detection can be integrated into the estimation model. In fact, the main objective of this chapter is to introduce the new concept of the FBMV and give some hints of how the information derived from the model itself can support the verification process. However, there are still open places to improve the process and sharpen its strength.

Chapter 7

Disscusion and Future Directions

7.1 Conclusion

Geometric modeling and description of 3D world objects collected through an imaging system –optical and/or electronical sensors– has become a topic of increasing importance, as they are essential for a variety of applications such as telecommunication, 3D city models, virtual tourist information system, etc. Inevitably, a fully human-based image interpretation system would be a costly and labour intensive operation. Therefore, there is an increasing demand towards fully automated machine-based image interpretation systems. This is a collection of processes that, to a varying degree, models the functionality of the human visual and cognitive system. This thesis addressed the problem of automatic recognition and 3D reconstruction of buildings from aerial images. It is mainly concerned with introducing the new concepts and development of robust methods in a hierarchical framework, for a data-driven reconstruction of generic plane-face building objects through the integration of computer vision and digital photogrammetric techniques. The term data-driven is used to indicate that the process of recognition and reconstruction is performed without a priori knowledge about building type or its structure, and the term generic is used to emphasize the fact that this type of reconstruction is not based on specific, user-defined building models, but rather on a generic one. Reconstruction based on a generic object model means that the number, as well as the geometric form, and the position of the significant parts of the model have to be defined. In addition, the geometric and topological relationships between these primitives are also needed. Finding the logical relationships between these geometric primitives when a specific object model is not present, is a complex problem, and its complexity is in a reciprocal-like relation with the geometrical level of the incorporated geometric primitives. That means, hypothesis model generation of a generic object based on point or line primitives is more complex than a polygonal-based approach. In addition, it is recognized that discontinuities should be represented by straight lines, and line segments can also be detected with sub-pixel accuracy and thus give a high quality result. The region-based segmentation algorithms may miss the relevant boundary information, and are generally unable to trace fine details and linear elements. They usually tend to produce regions of which the shape reflects more the search strategy used than the true shape of the regions. In consequence, an automated 3D image analysis should incorporate the descriptions of point, line, and region segments to admit a compact transfer of most of the information content in the image to higher level processes. This is the key issue and the strength of our proposed method. It enters into the high-level quantitative domain of the recognition process –extraction of regional information– in order to reduce the complexity of the problem in the very early stage of the whole chain of a generic-based reconstruction process, and integrates the low-level image-oriented qualitative geometric primitives –points and edges information–in the high-level model-oriented process during the hypothesis verification process. This is in contrary to most of the reported methods, which initialize their recognition process from the low-level geometric primitives, and are struggling with complex search strategies in the higher level processes.

An automated vision process such as 3D object reconstruction requires not only describing the geometry of the object of interest but also the ability to deal with incorrect data which will inevitably arise in a real system. They must be able to interpret the data while simultaneously reject the outliers. It has been shown when outliers contaminate the observations, the solution of the LS estimator becomes unreliable and it fails to correctly recover the model parameters. There are classes of computation in the field of robust statistics that have been designed to handle outliers. In this study we have surveyed the most common categories of robust parameter estimators. They were evaluated on relative efficiency and breakdown points. Robust methods based on random sampling techniques such as RANSAC, and LMS, are able to identify large fractions of outliers and perform well in presence of outliers but they are not optimal in suppressing Gaussian noise. Although outliers are a serious problem in vision and must be addressed in the formulation of vision algorithms, Gaussian noise is also present. M-estimators, which are more satisfying from a statistical standpoint, fail either the first initial estimate is too far away from the true solution, or the fraction of outliers goes beyond 35% of the data. The two-stage synthesis robust estimator proposed in this study is able to handle outliers and Gaussian error simultaneously and typically overcomes the problems. The combination of a random sampling type estimator to detect outliers, and estimate

the initial values of the parameters followed by an M-estimator, which is statistically more satisfying as it is a closer approximation to the maximum likelihood estimator is shown to be well suited for several parameter estimation problems which one may encounter in an automated vision process.

A new method for recognition of the 2D plane-roof regions, which possess meaningful correspondence to the 3D plane-roof polygons of the real world building roofs structure, is presented. The recognition is performed based on the pragmatic assumptions that roof structures are mostly planar surfaces, therefore introducing this information as two geometrically oriented constraints, 1) regional, and 2) planar geometric primitives, during segmentation process. The segmentation process incorporates the planarity knowledge based on the invariant geometric characteristics of surfaces, mean and Gaussian curvatures These two quantities provide a common method for specifying eight basic types of surfaces, e.g. flat, which are surronding any point on a smooth surface, thus yielding the coarse classification of surface types in image data. The regional constraint is integrated into the process using an iterative least squares planar fit region-growing procedure. Moreover, the required tuning parameters and thresholds for extraction of the primary roof structures are tied to the estimated noise variance of the corresponding image. The performance of the recognition procedure shows that the satisfactory results can be obtained using the proposed method. The intermediate results also indicate that the recognition of the microstructures of the building roof in the presence of high resolution image data is feasible. This type of information improves significantly the results of the subsequent reconstruction processes. In fact, the extracted 2D plane-roof regions are the basic elements to describe the geometry of the building during 3D reconstruction.

Modeling complex objects such as buildings requires considerable attention to their topology. We must understand how simple elements are connected to form the complex model, and how its topology is preserved when subjected to a variety of transformation. In the proposed framework of building reconstruction, the essential topological information such as 'adjacency' and 'contained-in' relationships between object primitives are provided by PAR (Polygons Adjacency Relationships). These are the minimum types of relationships between object primitives required in an automated vision process based on a generic object model. The PAR is defined based on Voronoi diagram in a raster domain, in such a way that shape and boundary of the polygons are also taken into account. In this manner the concept of spatial adjacency which has been normally defined based on a point-wise data set, is extended by introducing the adjacency relationships between polygonal primitives of different shapes and sizes, including connected, disconnected, or overlapped ones.

POLY-MODELER, a new mathematically founded model generator tool, is developed in this research study. It is a reliable and efficient tool for coarse polyhedral-like object model generation. The reconstruction is based on the 3D intersection of adjacent plane-roof polygons, and analysis of the intersection points. POLYMODELER determines where component faces are extended or truncated and new edges and vertices are created or deleted. When boundary elements overlap or coincide it merges them into a single element and thus maintains a consistent, non-redundant data structure representing model boundary. New edges are created where adjacent polygons intersect. The POLY-MODELER finds these intersections and then determines by point membership classification, which segments of the intersection are actual edges of the model. Assuming correctly oriented 3D plane-roof polygons along their adjacency relationships, the proposed method is able to recover the geometry of any generic plane-face building models. The performance of the method and accuracy of the reconstructed roof structure is highly related to the estimated parameters of the of the 3D plane-roof polygons. A failure in correctly recovering the surface normal of the 3D polygons causes an unexpected result leading to partially or completely wrong building description. This is why the quality of utilized DSM is of high importance in our approach. The results indicate that the recognition of microstructure on top of the building roof such as a dormer window is also possible. Nevertheless, in order to be able to estimate the correct pose of these microstructures on top of the roof and geometrically describe their shapes, a very dense and highly accurate DSM is required. Furthermore, owing to the geometrical reconstruction of roof structure, positional accuracy of roof elements such as orientation edges and intersection points are very high. However, due to a misinterpretation of the surface normal of polygonal primitives, we may have some discrepancies in the form of displacement or rotation from the real positions of these elements. To improve the quality of the generated model, the geometric and topological information provided by the coarse model is incorporated into a hypothesis model verification process. Since the model-driven geometric primitives (model edges) are used as the features of interest during matching process, the proposed method is classified as feature based model verification method (FBMV). The method is able to impose the geometric and topologic constraints derived from the initial descriptions of the object model into the estimation model, therefore increasing the reliability and accuracy of the estimation process. Furthermore, the ability to integrate all the available information to constrain the estimation of the model parameters significantly improves the model reconstruction. This capability is an essential component of an automated vision system. This is specially needed for automating the reconstruction of complex objects from real imagery. Practical experiments have proven that the intensity-based low-level vision processes solely could

not find both sufficient and relevant image features unless they incorporate the geometrically-based information into the process.

The FBMV is a general model verification method, and can be applied for fitting any solid object model with arbitrarily curved surfaces and with any number of model primitives to the homologous image features. In other words, its framework allows different, non-polyhedral object models to be used as well. A consequence of disregarding the above restriction on the geometry of the model primitives will necessitate that the geometric routine and functional model are adopted accordingly.

7.2 Directions for Further Research

As described in the previous section, this research has covered a number of issues in computer vision and photogrammetry, particularly recognition and 3D reconstruction of buildings. All the proposed methods have been implemented and tested, and it can be concluded that the research objectives set out in this thesis have been achieved. However, in a broader concept of creation of a true 3D geo-spatial information system, there are still issues that need to be investigated and studied such as those which are discussed in the introductory chapter. In addition, some of the aspects treated in this research also need further study and development which are summerized as follows:

- Identifying and developing new methods for extraction of regions of interest (ROI), with the particular emphasis in densely built-up areas. Possible lines of investigations are 1) integration of texture analysis and wavelet transformation, 2) image classification utilizing hight data as additional source of information and possibly analysing the shadow information (see section 4.4 for references).
- In this research the process of coarse reconstruction of hypothesis building model is performed using only one single image, therefore the results can improve significantly if the hypothesis model generation is performed in all the available correponding images in a parallel process, thus providing multiple hypothesis coarse models for every building object. The coarse building hypotheses can merge or fuse together in 3D object space in order to generate a unique and more reliable coarse building hypothesis to undergo the final verification process. In this manner, it is feasible to recover a part(s) of the building roof structure which is missing or has not been reconstructed in one or more of the generated coarse models because of e.g., occlusion, shadow or noise during segmentation process. It should be noted, although the FBMV method is capable of recovering most parts of the coarse model as it is also working based upon multiple images, however if the initial position of the geometric primitive(s) of the coarse building structure is far away from its real position, the FBMV will fail to modify that part(s) of the building structure. Applying the ICP (Iterative Closest Point) algorithm for the registration of the 3D shapes reported by (Besl & McKay 1992) and extened by (Gühring 1999) to merge coresponding 3D object models would be a starting point to tackle this kind of problem.
- Further testing and improving the FBMV algorithm by applying robust parameter estimation techniques such as M-estimator discussed in chapter 3 for detection of outliers, identyfying and integrating the remaining model driven constraints into the estimation process, and proposing the new fields of applications for the Feature Based Model Verification concept such as interactive or semi-autmatic object reconstruction, or its application in the industrial domain.

I hope that the proposed method in this study in the long run forms the fundamental basis for automating the process of modeling the real world complex objects while ensuring the consistency, accuracy and reliability of the reconstructed model.

Bibliography

Ackermann, F. (1984), 'Digital image correlation: Performance and potential application in photogrammetry', *Photogrammetric Record* **11**(64), 429–439.

Ackermann, F. & Krzystek, P. (1991), 'MATCH-T: Automatic mensuration of digital elevation models', Presented paper to the 3rd Technical Seminar of the Sociedad Espanola de Cartografia Fotogrametria y Teledeteccion.

Ameri, B. (2000), Feature Based Model Verification (FBMV): a new concept for hypothesis validation in building reconstruction, *in* 'IAPRS, Vol XXXIII, Part 3', Amsterdam, the Netherlands.

Ameri, B. & Fritsch, D. (1999), 3-D reconstruction of polyhedral-like building models, *in* 'IAPRS, Vol 32, Part 3-2W5', Munich, Germany, pp. 15–20.

Ameri, B. & Fritsch, D. (2000), Automatic 3D building reconstruction using plane-roof structures, *in* 'Proc. of ASPRS Annual Conference', Washington DC, USA.

Axelsson, P. (1996), Autonomous Decisions in Photogrammetry using Minimum Description Length, PhD thesis, Royal Institute of Technology, Dept. of Geodesy and Photogrammetry, Stockholm, Sweden.

Axelsson, P. (1997), Interactive 3D extension of 2D map data using mono images, *in* 'ISPRS Workshop on 3D Reconstruction and Modelling of Topographic Objects', Stuttgart, Germany, pp. 145–150.

Baarda, W. (1967), 'Statistical concepts in geodesy', *Netherlands Geod. Commission New Series* **2**(4).

Baarda, W. (1968), 'A testing procedure for use in geodetic networks', *Netherlands Geod. Commission New Series* **2**(5).

Baillard, C., Schmid, C., Zisserman, A. & Fitzgibbon, A. (1999), Automatic line matching and 3D reconstruction of buildings from multiple views, *in* 'IAPRS, Vol 32, Part 3-2W5', Munich, Germany, pp. 69–80.

Baltsavias, E., Eckstein, W., Gülch, E., Hahn, M., Stallmann, D., Tempfli, K. & Welch, R., eds (1997), *3D Reconstruction and Modelling of Topographic Objects*, Vol. 32, B3-4W2, ISPRS, Stuttgart, Germany.

Baltsavias, E., Mason, S. & Stallmann, D. (1995), Use of DTMs/DSMs and orthoimages to support building extraction, *in* A. Gruen, O. Kuebler & P. Agouris, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 199–210.

Baltsavias, E. P. (1991), Multiphoto Geometrically Constrained Matching, PhD thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland.

Baltsavias, E. P. (1999), 'A comparison between photogrammetry and laser scanning', *ISPRS Journal* **54**(2-3), 83–94.

Besl, P. J. (1986), Surfaces in Early Range Image Understanding, PhD thesis, University of Michigan, Ann Arbor, Rep. RSD-TR-10-86.

Besl, P. J. & Jain, R. C. (1988), 'Segmentation through variable order surface fitting', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**(2), 167–192.

Besl, P. J. & McKay, N. D. (1992), 'A method for registration of 3-d shapes', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(2), 239–256.

Best, M. J. & Ritter, K. (1985), *Linear Programming: Active Set Analysis and Computer Programs*, Prentice-Hall Inc., Englewood Cliffs, NJ.

Bignone, F., Henricson, O., Fua, P. & Stricker, M. (1996), Automatic extraction of generic house roofs from high resolution aerial imagery, *in* 'Computer Vision - ECCV96', pp. 85–96.

Bolle, R. M. & Cooper, D. B. (1984), 'Bayesian recognition of local 3-D shape by approximating image intensity functions with quadric polynomials', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(4), 418–429.

Bolles, R. C. & Fischler, M. A. (1981), A ransac-based approach to model fitting and its application to finding cylinders in range data, *in* 'Seventh Int. Joint Conference on Artificial Intelligence', pp. 637–643.

Borgefors, G. (1986), 'Distance transformation in digital images', *Computer Vision, Geraphics, and Image Processing* **34**, 344–371.

Braun, C., Kolbe, T., Lang, F., Schickler, W., Steinhage, V., Kremers, A., Förstner, W. & Plümmer, L. (1995), 'Models for photogrammetric building reconstruction', *Computer & Graphics* **19**(1), 109–118.

Brenner, C. (1999), Interactive modelling tools for 3D building reconstruction, *in* D. Fritsch & R. Spiller, eds, 'Photogrammetric Week '99', Herbert Wichmann Verlag, Heidelberg, pp. 23–34.

Brenner, C. & Haala, N. (1998*a*), Fast production of virtual reality city models, *in* 'IAPRS, Vol 32, Part 4', Stuttgart, Germany.

Brenner, C. & Haala, N. (1998*b*), Rapid acquisition of virtual reality city models from multiple data sources, *in* 'IAPRS, Vol 32, Part 5', Hakodate, Japan.

Brooks, R. A. (1981), 'Symbolic reasoning among 3-D models and 2-D images', *Artificial Intelligence* **17**, 1285–348.

Brunn, A., Lang, F. & Förstner, W. (1996), A procedure for segmenting surfaces by symbolic and iconic image fusion, *in* 'Mustererkennung 96, Proceeding of the DAGM 96', Springer-Verlag, Heidelberg, pp. 11–20.

Brunn, A. & Weidner, U. (1997), Extracting buildings from digital surface models, *in* 'ISPRS Workshop on 3D Reconstruction and Modelling of Topographic Objects', Stuttgart, Germany, pp. 27–34.

Canny, J. (1986), 'A computational approach to edge detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**, 679–698.

Chaterjee, S. & Hadi, A. S. (1988), *Sensitivity Analysis in Linear Regression*, John Wiley, New York.

Collins, R., Hanson, A., Riseman, E. & Schultz, H. (1995), Automatic extraction of buildings and terrain from aerial images, *in* A. Gruen, O. Kuebler & P. Agouris, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 169–178.

CVIU (1998), 'Special issue on Building Extraction', *Computer Vision and Image Understanding* **72**(2).

Danahy, J. (1999), Visualization data needs in urban environmental planning and design, *in* D. Fritsch & R. Spiller, eds, 'Photogrammetric Week '99', Herbert Wichmann Verlag, Heidelberg, pp. 351–365.

Duperet, A., Eidenbenz, C. & Holland, D. (1997), Capturing and maintaining datasets using digital imagery: Experience and future requirements of national mapping agencies, *in* 'ISPRS Workshop on 3D Reconstruction and Modelling of Topographic Objects', Vol. 32, part 3-4W2, Stuttgart, Germany, pp. 51–67.

Ebner, E., Eckstein, W., Heipke, C. & Mayer, H., eds (1999), *Automatic Extraction of GIS Objects from Digital Imagery*, Vol. 32, B3-2W5, ISPRS, Munich, Germany.

Eckstein, W. & Munkelt, O. (1995), Extracting objects from digital terrain models, *in* 'in T. Schenk (ed), 'Remote Sensing and Reconstruction for Three-Dimensional Objects and Scenes, SPIE Vol. 2572', San Diego, California, pp. 43–51.

Englert, R. (1997), Systematic acquisition of generic 3D building model knowledge, *in* 'Workshop on Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, SMATI'97', pp. 181–195.

Fatemi Ghomi, N. (1997), Performance Measures for Wavelet-based Segmentation Algorithms, PhD thesis, Center for Vision, Speech and Signal Processing, University of Surrey.

Faugeras, O. (1996), *Three-Dimensional Computer Vision, A Geometric Viewpoint*, The MIT Press, Cambridge.

Faugeras, O., Laveau, S. & Robert, L. (1995), 3D reconstruction of urban scenes from sequences of images, *in* A. Gruen, O. Kuebler & P. Agouris, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 145–168.

Fischer, A., Kolbe, T. H. & Lang, F. (1999), On the use of geometric and semantic models for component-based building reconstructoin, *in* 'Workshop on Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, SMATI'99', pp. 101–119.

Fischer, A., Kolbe, T., Lang, F., Cremers, A. B., Förstner, W., Plümer, L. & Steinhage, V. (1998), 'Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D', *Computer Vision and Image Understanding* **72**(2), 195–203.

Fischler, M. A. & Bolles, R. C. (1981), 'Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography', *Communications of the ACM* **24, No. 6**, 381–395.

Förstner, W. (1982), On the geometric precision of digital correlation, *in* 'International Archives of Photogrammetry and Remote Sensing', Vol. 24, 3/3, pp. 176–189.

Förstner, W. (1986), A feature based correspondence algorithm for image matching, *in* 'International Archives of Photogrammetry and Remote Sensing', Vol. 26, 3/3, Rovaniemi.

Förstner, W. (1987), 'Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision', *Computer Vision, Graphics, and Image Processing* **40**, 273–310.

Förstner, W. (1989), Image analysis techniques for digital photogrammetry, *in* 'Proceedings of the 42th Photogrammetric Week', Schriftenreihe des Institus f"ur Photogrammetrie der Universit"at Stuttgart, Heft 13, pp. 205–221.

Förstner, W. (1995), Mid-level vision processes for automatic building extraction, *in* A. Gruen, O. Kuebler & P. Agouris, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 179–188.

Förstner, W. (1996), 10 pros and cons against performance characterization of vision algorithms, *in* 'Workshop on Performance Characterization of Vision Algorithms', Cambridge.

Förstner, W. (1999), 3D-city models: Automatic and semiautomatic acquisition methods, *in* D. Fritsch & R. Spiller, eds, 'Photogrammetric Week '99', Herbert Wichmann Verlag, Heidelberg, pp. 291–303.

Förstner, W., Liedtke, C. E. & Bückner, J., eds (1999), *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, Institut fr Photogrammetrie, Universitt Bonn, Germany.

Förstner, W. & Plümer, L., eds (1997), *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, Birkhäuser Verlag, Bonn, Germany.

Fradkin, M., Roux, M. & Maitre, H. (1999), Building detection from multiple views, *in* 'IAPRS, Vol 32, Part 3-2W5', Munich, Germany, pp. 81–86.

Fritsch, D. (1985), Some additional informations on the capacity of the linear complementary algorithm, *in* E. W. Grafarend & F. Sanso, eds, 'Optimization and Design of Geodetic Networks', Springer-Verlag, Heidelberg, pp. 169–184.

Fritsch, D. (1999), Virtual cities and landscape models-what has photogrammetry to offer?, *in* D. Fritsch & R. Spiller, eds, 'Photogrammetric Week '99', Herbert Wichmann Verlag, Heidelberg, pp. 3–14.

Fritsch, D. & Ameri, B. (1998), Geometric characteristics of digital surfaces: A key towards 3D building reconstruction, *in* 'IAPRS, Vol 32, Part 3/1', Columbus, Ohio, USA, pp. 119–126.

Fryer, M. J. (1978), *An Introduction to Linear Programming and Matrix Game Theory*, Edward Arnold, London.

Fua, P. (1996), Model-based optimization: Accurate and consistent modeling, *in* 'IAPRS, Vol XXXI, Part 3', Wien, pp. 222–233.

Fua, P. & Hanson, A. (1987), 'Resegmentation using generic shape: Locating general cultural objects', *Pattern Recognition Letters* **5**(3), 243–252.

Fua, P. & Hanson, A. (1988), Extracting generic shapes using model-driven optimization, *in* 'DAPRA Image Understanding Workshop', Morgan Kaufmann Publishers, Cambridge, Massachusetts, pp. 994–1004.

Fua, P. & Leclerc, Y. G. (1990), 'Model driven edge detection', *Machine Vision and Applications* (3), 45–56.

Fuchs, C., Gülch, E. & Förstner, W. (1998), *OEEPE Survey on 3D-City Models, OEEPE Publication No. 35*, Bundesamt für Kartographie und Geodäsie, Frankfurt.

Gold, C., M. (1990), Neighbours, adjacency and theft - the voronoi process for spatial analysis, *in* 'Proceedings, First European conference on Geographic Information Systems', Amsterdam, pp. 382–391.

Gonzalez, R. & Woods, R. (1993), *Digital Image Processing*, Addison-Wesley Publishing, Reading, Massachusetts.

Gruber, M. (1999), Managing large 3D urban databases, *in* D. Fritsch & R. Spiller, eds, 'Photogrammetric Week '99', Herbert Wichmann Verlag, Heidelberg, pp. 341–349.

Gruber, M., Pasko, M. & Leberl, F. (1995), Geometric versus texture deatail in 3-D models of real world buildings, *in* A. Gruen, O. Kuebler & P. Agouris, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 189–198.

Gruen, A. (1999), Cybercity modeler, a tool for interactive 3-D city model generation, *in* D. Fritsch & R. Spiller, eds, 'Photogrammetric Week '99', Herbert Wichmann Verlag, Heidelberg, pp. 317–327.

Gruen, A., Baltsavias, E. & Henricson, O., eds (1997), *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Vol. II, Birkhäuser Verlag, Basel, Boston, Berlin.

Gruen, A. & Dan, H. (1997), Tobago - a topology builder for the automated generation of building models, *in* A. Gruen, O. Kuebler & M. Baltsavias, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 149–160.

Gruen, A., Kuebler, O. & Agouris, P., eds (1995), *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Vol. I, Birkhäuser Verlag, Basel, Boston, Berlin.

Gruen, A. & Li, H. (1997), Linear feature extraction with 3-D lsb-snakes, *in* A. Gruen, O. Kuebler & M. Baltsavias, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 287–298.

Gruen, A. & Stallmann, D. (1991), High accuracy edge matching with an extension of the mpgc-matching algorithm, *in* 'Industrial Vision Metrology, SPIE Proceedings Series', Vol. 1526, pp. 42–55.

Gühring, J. (1999), *Integration of Accuracy Information into ICP*, Internal report, Institute for Photogrammetry (ifp), University of Stuttgart.

Gülch, E., Müller, H. & Läbe, T. (1999), Integration of automatic processes into semi-automatic building extraction, *in* 'IAPRS, Vol 32, Part 3-2W5', Munich, Germany, pp. 177–186.

Gülch, E., Müller, H., Läbe, T. & Ragia, L. (1998), 3-D reconstruction of polyhedral-like building models, *in* 'IAPRS, Vol 32, Part 3/1', Columbus, Ohio, pp. 331–338.

Haala, N. (1995), 3D building reconstruction using linear edge segments, *in* D. Fritsch & D. Hobbie, eds, 'Photogrammetric Week '95', Herbert Wichmann Verlag, Heidelberg, pp. 19–28.

Haala, N. (1996), Gebäuderekonstruktion durch Kombination von Bild- und Höhendaten, PhD thesis, Deutsche Geodätische Kommission, München, Vol. C 460.

Haala, N. (1999), Combining multiple data sources for urban data acquisition, *in* D. Fritsch & R. Spiller, eds, 'Photogrammetric Week '99', Herbert Wichmann Verlag, Heidelberg, pp. 329–339.

Haala, N. & Anders, K.-H. (1996), Fusion of 2D-GIS and image data for 3D building reconstruction, *in* 'IAPRS, Vol XXXI, Part 3', Wien, pp. 285–290.

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

Haala, N. & Brenner, C. (1999), 'Extraction of buildings and trees in urban environments', *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(2-3), 130–137.

Haala, N., Brenner, C. & Anders, K.-H. (1998), 3D Urban GIS from Laser Altimeter and 2D Map Data, *in* 'IAPRS, Vol 32, Part 3/1', Columbus, Ohio, USA, pp. 339–346.

Haala, N. & Vosselman, G. (1992), Recognition of road and river patterns by relational matching, *in* 'Proc. ISPRS Congress Comm. III', Washington D.C., pp. 969–975.

Hampel, F. R. (1968), Contributions to the theory of robust estimation, PhD thesis, University of California, Berkeley.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986), *Robust Statistics*, John Wiley and Sons, New York.

Haralick, R. M. (1984), 'Digital step edges from zero-crossing of second directional derivatives', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(1), 58–68.

Haralick, R. M. & Joo, H. (1988), 2D-3D pose estimation, *in* '9th Int. Conference on Pattern Recognition', Rome, Italy, pp. 385–391.

Haralick, R. M., Shanmugam, K. & Dinstein, I. (1973), 'Textural features for image classification', *IEEE Transactions on System, Man, and Cybernetics* **SMC**(3), 610–621.

Haralick, R. & Shapiro, L. (1992), *Computer and Robot Vision*, Vol. 2, Addison-Wesley Publishing Company.

Hendrickx, M., Vandekerckhove, J., Frere, D., Moons, T. & Gool, L. C. (1997), 3D reconstruction of house roofs from multiple aerial images of urban areas, *in* 'ISPRS Workshop on 3D Reconstruction and Modelling of Topographic Objects', Stuttgart, Germany, pp. 88–95.

Henricson, O. (1998), 'The role of color attributes and similarity grouping in 3-D building reconstruction', *Computer Vision and Image Understanding* **72**(2), 163–184.

Henricson, O., Bignone, F., Willhuhn, W. & Ade, F. (1996), Project Amobe: Strategies, current status and future work, *in* 'IAPRS, Vol XXXI, Part 3', Wien, pp. 321–330.

Heuser, M. & Liedtke, C. (1990), Recognition of the spatial position of industrial 3d-objects using relaxation techniques, *in* 'ICPR90', Vol. I, pp. 191–193.

Huber, P. J. (1981), *Robust Statistics*, John Wiley and Sons, New York.

Huertas, A., Lin, C. & Nevatia, R. (1993), Detection of buildings from monocular views of aerial scenes using perceptual organization and shadows, *in* 'in Proceeding of the 1993 ARPA Image Understanding Workshop', pp. 253–260.

Huertas, A. & Nevatia, R. (1988), 'Detecting buildings in aerial images', *Computer Vision, Graphics and Image Processing* **41**(2), 131–152.

Irvin, R. & McKeown, D. (1989), 'Methods for exploiting the relationship between buildings and their shadows in aerial imagery', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(6), 1564–1575.

Jaynes, C., Hanson, A. & Riseman, E. (1997), Model-based surface recovery of building in optical and range images, *in* 'Workshop on Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, SMATI'97', pp. 212–227.

Kim, D. Y., Kim, J. J., Meer, P., Mintz, D. & Rosenfeld, A. (1989), Robust computer vision: A least median of squares based approach, *in* 'Proc. of the DARPA Image Understanding Workshop', Palo Alto, CA, USA.

Kim, T. & Mueller, J. P. (1995), Building extraction and verfification from spaceborne and aerial imagery using image understanding fusion techniques, *in* A. Gruen, O. Kuebler & P. Agouris, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 221–230.

Koch, K. R. (1999), *Parameter Estimation and Hypothesis Testing in Linear Models*, Springer-Verlag, Berlin Heidelberg.

Kofler, M., Rehatschek, H. & Gruber, M. (1996), A database for a 3D gis for urban environments supporting photorealistic visualization, *in* 'Proceedings of the ISPRS Congress, Part B2', Wien, pp. 198–202.

Kulschewski, K. & Koch, K. R. (1997), Building recognition with bayesian networks, *in* 'Workshop on Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, SMATI'97', pp. 196–210.

Kulschewski, K. & Koch, K. R. (1999), Recognition of buildings using a dynamic bayesian network, *in* 'Workshop on Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, SMATI'99', pp. 121–132.

Kumar, R. & Hanson, A. R. (1990), Analysis of different robust methods for pose refinement, *in* 'Int. Workshop on Robust Computer Vision', Seattle, USA, pp. 167–182.

Lang, F. (1999), Geometrische und semantische Rekonstruktion von Gebäuden durch Ableitung von 3D-Gebädeecken, PhD thesis, Institute of Photogrammetry, University Bonn.

Lang, F. & Förstner, W. (1996), 3D-City modeling with a digital one-eye stereo system, *in* 'IAPRS, Vol XXXI, Part 4', Wien.

Lange, E. (1999), The degree of realism of gis-based virtual landscapes: Implications for spatial planing, *in* D. Fritsch & R. Spiller, eds, 'Photogrammetric Week '99', Herbert Wichmann Verlag, Heidelberg, pp. 367–374.

Leberl, F., Walcher, W., Wilson, R. & Gruber, M. (1999), Models of urban areas for line-of-sight analyses, *in* 'IAPRS, Vol 32, Part 3-2W5', Munich, Germany, pp. 217–226.

Lee, C. N., Haralick, R. M. & Zhuang, X. (1989), Recovering 3-D motion parameters from image sequences with gross errors, *in* 'IEEE Workshop on Visual Motion', Irvine, USA.

Lee, D. & Schenk, T. (1992), Image segmentation from texture measurement, *in* 'IAPRS, Comm. III', Washington D.C., USA, pp. 195–199.

Lee, D. & Schenk, T. (1998), An adaptive approach for extracting texture information and segmentation, *in* 'IAPRS, Vol 32, Part 3/1', Columbus, Ohio, USA, pp. 250–255.

Lemmens, M. J. P. M. (1996), Structure-Based Edge Detection, PhD thesis, Technical University of Delft.

Lin, C., Huertas, A. & Nevatia, R. (1995), Detection of buildings from monocular images, *in* A. Gruen, O. Kuebler & P. Agouris, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 125–134.

Lohr, U. & Eibert, M. (1995), The TopoSys laser scanner-system, *in* D. Fritsch & D. Hobbie, eds, 'Photogrammetric Week '95', Herbert Wichmann Verlag, Heidelberg, pp. 263–268.

Lowe, D. (1985), *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, Boston, Mass.

Lowe, D. G. (1991), 'Fitting parameterized three-dimensional models to images', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(5), 441–450.

Maas, H. G. (1999), Closed solutions for the determination of parametric building models from invariant moments of airborne laserscanner data, *in* 'IAPRS, Vol 32, Part 3-2W5', Munich, Germany, pp. 193–199.

Mason, S., Baltsavias, M. & Stallmann, D. (1994), High precision photogrammetric data set for building reconstruction and terrain modelling, Data description, Institute for Geodesy and Photogrammetry, ETH Zurich.

McKeown, D. & McGlone, J. (1993), Integration of photogrammetric cues into cartographic feature extraction, *in* 'SPIE Conference on Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision', Orlando, Florida, pp. 2–15.

Midtbø, T. (1993), Spatial Modelling by Delaunay Networks of Two and Three Dimensions, PhD thesis, Norwegian Institute of Technology, University of Tronheim.

Mikhail, E. M. (1976), *Observations and Least Squares*, IEP-A Dun-Donnelley Publisher, New York.

Mohan, R. & Nevatia, R. (1989), Segmentation and description based on perceptual organization, *in* 'Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition', San Diego, CA, pp. 333–341.

Mortenson, M. E. (1997), *Geometric Modeling*, 2 edn, John Wiley and Sons, New York.

Mueller, W. & Olson, J. (1993), Model-based feature extraction, *in* 'SPIE Conference on Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision', Orlando, Florida, pp. 263–272.

Nagao, M. & Matsuyama, T. (1980), *A Structural Analysis of Complex Aerial Photographs*, Plenum Press, New York.

Nebiker, S. & Carosio, A. (1995), Automatic extraction and structuring of objects from scanned topographic maps- an alternative to the extraction from aerial and space images, *in* A. Gruen, O. Kuebler & P. Agouris, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 287–2296.

Nevatia, R., Huertas, A. & Kim, Z. (1999), The muri projects for rapid feature extraction in urban areas, *in* 'IAPRS, Vol 32, Part 3-2W5', Munich, Germany, pp. 3–14.

Nevatia, R., Lin, C. & Huertas, A. (1997), A system for building detection from aerial images, *in* A. Gruen, O. Kuebler & M. Baltsavias, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 77–86.

Oddo, L. A. (1992), Global shape entropy: a mathematically tractable approach to building extraction in aerial imagery, *in* '20th AIPR Workshop, Computer Vision Applications: Meeting the challenges, SPIE Vol. 1623', pp. 91–101.

Pilouk, M., Tempfli, K. & Molenaar, M. (1994), A tetrahedron-based 3D vector data model for geoinformation, *in* 'In Proc. Advanced geographic data modelling: spatial data modelling and query languages for 2D and 3D applications', Delf, The Netherlands, pp. 129–140.

Pitas, I. (1993), *Digital Image Processing Algorithms*, Prentice Hall International, UK.

Preparata, F. P. & Shamos, M. I. (1985), *Computational Geometry*, Springer-Verlag, New York.

Price, K. & Huertas, A. (1992), Using perceptual grouping to detect objects in aerial scenes, *in* 'Proc. ISPRS Congress Comm. III', Washington D.C., pp. 842–855.

Rissanen, J. (1987), 'Minimum description length principle', *Encyclopedia of Statistical Sciences* **5**, 523–527.

Roth, G. & Levine, M. D. (1990), Random sampling for primitive extraction, *in* 'Int. Workshop on Robust Computer Vision', Seattle, USA, pp. 352–366.

Rousseeuw, P. J. & Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, John Wiley and Sons, New York.

Sagerer, G., Kummert, F. & Socher, G. (1996), Semantic models and object recognition in computer vision, *in* 'Proc. ISPRS Congress Comm. III', Vol. XXXI/B3, Vienna, pp. 710–723.

Salesh, S. & Sowmya, A. (1998), Rall: Road recognition from aerial images using inductive learning, *in* 'IAPRS, Vol 32, Part 3/1', Columbus, Ohio, USA, pp. 367–378.

Sali, E. & Wolfson, H. (1992), 'Texture classification in aerial photographs and satellite data', *Int. Journal Remote Sensing* **13**(18), 3395–3408.

Schenk, T. (1993), Image understanding and digital photogrammetry, *in* D. Fritsch & D. Hobbie, eds, 'Photogrammetric Week '93', Herbert Wichmann Verlag, Heidelberg, pp. 197–207.

Schenk, T. & Toth, C. (1991), Reconstructing visible surfaces, *in* 'SPIE Proceedings of Industrial Vision Metrology', Vol. 1526, Winnipeg, Canada, pp. 78–89.

Schickler, W. (1992), Feature matching for outer orientation of single images using 3-D wireframe controlpoints, *in* 'Proc. ISPRS Congress Comm. III', Washington DC, USA, pp. 591–598.

Schmid, C. & Zisserman, A. (1997), Automatic line matching across views, *in* 'Proc. CVPR', pp. 666–671.

Schunck, B. G. (1990), Robust computational vision, *in* 'Int. Workshop on Robust Computer Vision', Seattle, USA, pp. 1–18.

Schutte, K. & Hilhorst, G. H. J. (1993), Comparison levels for iterative estimators for model-based recognition of man-made objects in remote sensing images, *in* 'Proc. SPIE Symposium on Electronic Imaging: Science and Technology', Vol. 1904, San Jose, pp. 222–228.

Sester, M. & Förstner, W. (1989), Object location based on uncertain models, *in* 'Mustererkennung 1989', Vol. 219 of *Informatik Fachberichte*, Springer Verlag, Hamburg, pp. 457–464.

Shi, Z., Shibasaki, R. & Murai, S. (1997), Automatic building extraction from digital stereo imagery, *in* A. Gruen, O. Kuebler & M. Baltsavias, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 119–128.

Shufelt, J. & McKeown, D. (1993), 'Fusion of monocular cues to detect man-made structures in aerial images', *CVGIP: Image Understanding* **57**(3), 307–330.

Siebe, E. & Büning, U. (1997), Application of digital photogrammetric products for cellular radio network planning, *in* D. Fritsch & D. Hobbie, eds, 'Photogrammetric Week '97', Herbert Wichmann Verlag, Heidelberg, pp. 159–164.

Sinha, S. S. & Schunck, B. G. (1989), A two-stage algorithm for discontinuity-preserving surface reconstruction, *in* 'Workshop on Robust Estimation', Maryland, USA.

Spreeuwers, L., Schutte, K. & Houkes, Z. (1997), A model driven approach to extract building from multi-view aerial imagery, *in* A. Gruen, O. Kuebler & M. Baltsavias, eds, 'Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)', Birkhäuser Verlag, Basel, Boston, Berlin, pp. 109–118.

Stilla, U., Geibel, R. & Jurkiewicz, K. (1997), Building reconstruction using different views and context knowledge, *in* 'ISPRS Workshop on 3D Reconstruction and Modelling of Topographic Objects', Stuttgart, Germany, pp. 129–136.

Strat, T. (1994), Photogrammetry and knowledge representation in computer vision, *in* 'IAPRS, Vol 30, Part 3', Munich, Germany, pp. 784–792.

Strat, T. & Fischler, M. (1991), 'Context-based vision: Recognizing objects using information from both 2-D and 3-D imagery', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(10), 1050–1065.

Tang, L. (1992), Raster algorithms for surface modelling, *in* 'Proc. ISPRS Congress Comm. III', Washington D.C., pp. 566–573.

Torr, P. H. S. & Murray, D. W. (1993), Outlier detection and motion segmentation, *in* 'Proc. Sensor Fusion VI', SPIE volume 2059, Boston MA, pp. 432–443.

Torr, P. H. S. & Zisserman, A. (1997), 'Robust parametrization and computation of the trifocal tensor', *Image and Vision Computing* **15**, 591–607.

Tou, J. T. & Gonzalez, R. C. (1974), *Pattern Recognition Principles*, Addison–Wesley, London.

Volz, S. & Klinec, D. (1999), Nexus: The development of a platform for location aware application, *in* 'Proceedings of the third Turkish-German joint Geodetic Days', Vol. 2, Istanbul, Turkey, pp. 599–608.

Vosselman, G. (1999), Building reconstruction using planar faces in very high density height data, *in* 'IAPRS, Vol 32, Part 3-2W5', Munich, Germany, pp. 87–92.

Waegli, B. (1998), Investigation into the noise characteristics of digital aerial images, *in* 'IAPRS, Vol 32, Part 2', Cambridge, UK, pp. 341–348.

Walter, V. (1999), Automated gis data collection and update, *in* D. Fritsch & R. Spiller, eds, 'Photogrammetric Week '99', Herbert Wichmann Verlag, Heidelberg, pp. 267–280.

Wang, Z. & Schenk, T. (1992), 3D urban area surface analysis, *in* 'IAPRS XVII, Vol 29, B3', pp. 720–726.

Wehr, A. & Lohr, U. (1999), 'Airborne laser scanning–an introduction and overview', *ISPRS Journal* **54**(2-3), 68–82.

Weidner, U. (1996), 'An approach to building extraction from digital surface models', pp. 924– 929.

Weidner, U. & Förstner, W. (1995), 'Towards automatic building extraction from high resolution digital elevation models', *ISPRS Journal* **50**(4), 38–49.

Wild, D., Krzystek, P. & Madani, M. (1996), Automatic breakline detection using an edge preserving filter, *in* 'IAPRS, Vol XXXI, Part 3', Wien, pp. 946–952.

Wiman, H. & Axelsson, P. (1996), Finding 3D-structures in multiple aerial images using lines and regions, *in* 'Proc. ISPRS Congress Comm. III', Wien, pp. 953–959.

Wouwer, G. V. (1998), Wavelets for Multiscale Texture Analysis, PhD thesis, University of Antwerpen.

Zhang, S., Sullivan, G. & Baker, K. (1992), 'Relational model construction and 3D object recognition from single 2D monochromatic image', *Image and Vision Computing* **10**(5), 313–318.

Zong, J., Li, J.-C. & Schenk, T. (1992), Aerial image matching based on zero crossing, *in* 'Proc. ISPRS Congress Comm. III', Washington D.C.

Appendix A

Experimental Results

This appendix portrays the results of the proposed methods for the coarse polyhedral-like model generation, and the final verification process, applied to the residential test images of the international Avenches data set (Mason et al. 1994). The types and the specifications of the data which has been used in this study are as follows:

- Digital stereo aerial images of scale $1 : 5.000$ with $15\mu_m$ resolution. The images captured with 60 % forward and sidewards overlap.
- Accurate orientation parameters.
- A dense DSM computed by commercial photogrammetric software with $0.25m$ ground resolution.

Figure (A.1), depicts the result of the reconstructed coarse buildings obtained by POLY-MODELER, overlaid on the correponding aerial image. Figure (A.2-b), shows a perspective view of the reconstructed coarse buildings in 3D object space. All the buildings are coarsely reconstructed, even building no. 4, which was under construction during aerial photography, is somehow reconstructed, which shows the robustness of the reconstruction process. In fact, the generated model is not accurate enough to be verified automatically applying FBMV, however the modification process is capable of sending a warning signal to the operator in order to check the final result and, if necessary, edit the model manually. This warning message is triggered based on the analysis of the estimated variances of the model points discussed in section 6.6. Figure (A.3), represents the verified fine reconstructed buildings obtained by FBMV technique, overlaid on the correponding aerial image. Figure (A.4), shows a perspective view of the final buildings in 3D object space. The results reveal that the FBMV is precisely modified and recovered the details of all the buildings except buildings no. 1, and 4. The problem with building no. 4 is already discussed above. The roof part of the building no. 1 has not been detected during the segmentation process. This is due to the occlusion of this part of the roof by the adjacent tree. As discussed in chapter 7, the problem of occlusion can be significantly eliminated by generating multiple coarse hypothesis models using multiple overlapping aerial images. The figure (A.5), and (A.6), show the perspective views of 3D reconstructed coarse and fine buildings ovelaid on the existing DSM respectively.

Figure A.1: Reconstructed coarse buildings overlaid on the corresponding aerial image

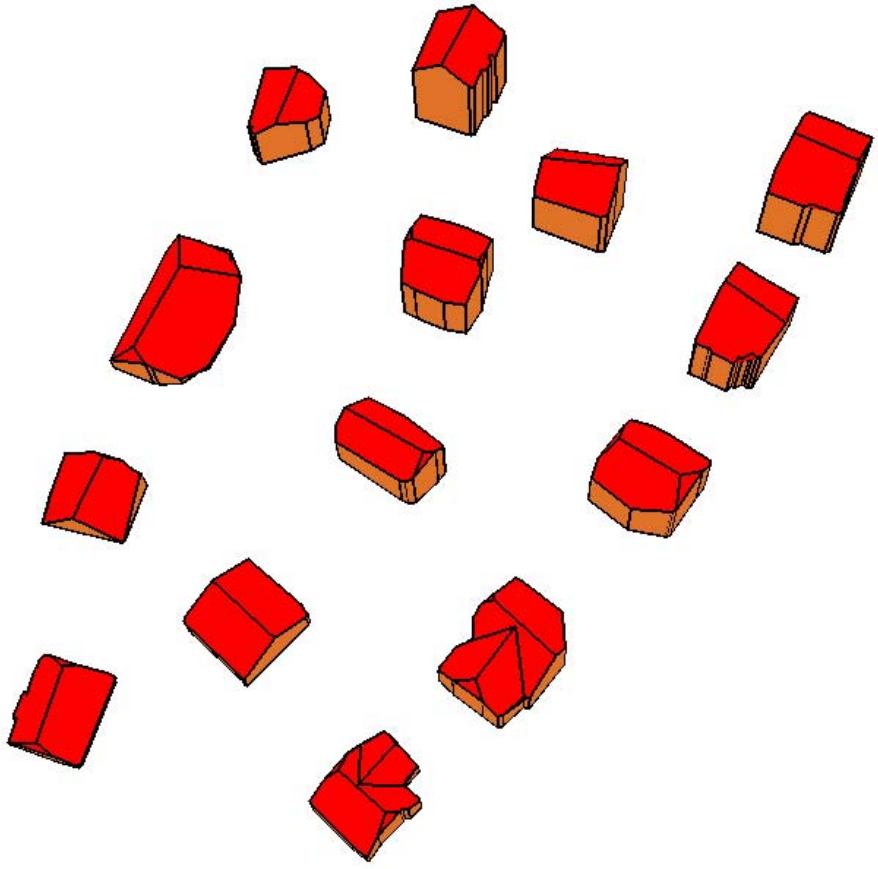


Figure A.2: Perspective view of 3D reconstructed coarse buildings

Figure A.3: final reconstructed buildings overlaid on the corresponding aerial image

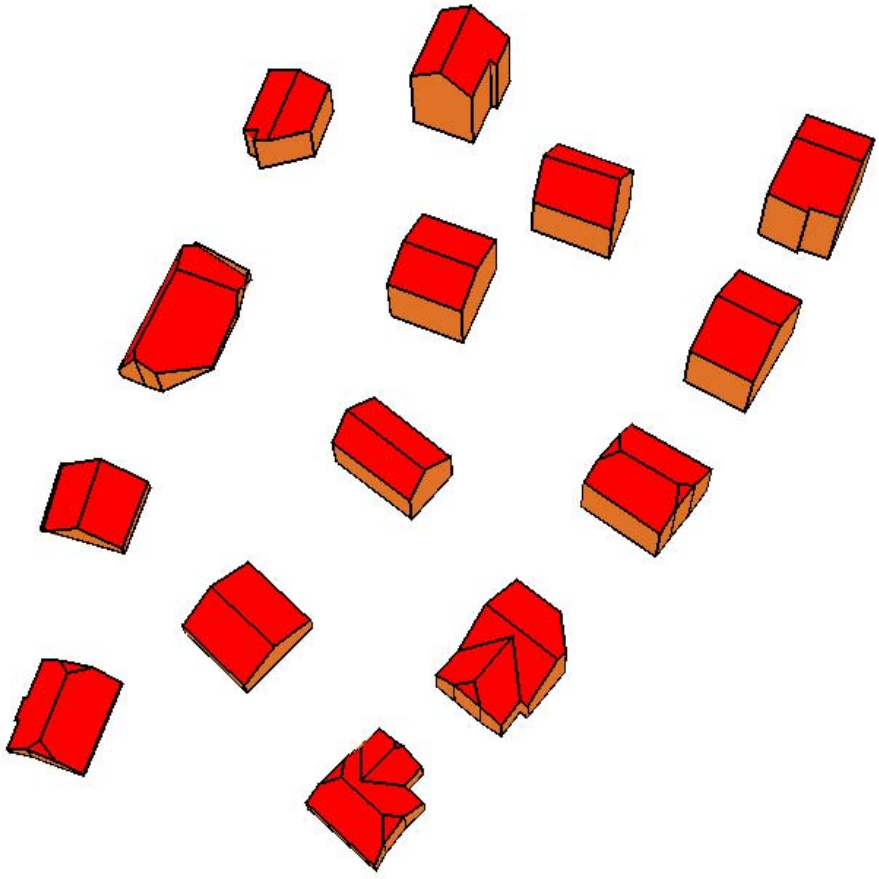


Figure A.4: Perspective view of the final 3D reconstructed buildings

Figure A.5: Perspective view of 3D reconstructed coarse buildings overlaid on the DSM

Figure A.6: Perspective view of the final 3D reconstructed buildings overlaid on the DSM

Acknowledgements

This thesis would not have been realized without the assistance, guidance and support of many people, all of whom it is not possible to mention by name.

I would like to express my sincere gratitude to Professor Dieter Fritsch for his kind supervision and many good ideas. His calcualted blending of guidance and freedom has not only made this study a possibility, but prepared me as a research trainee. It is hard to express how much I am obliged.

I am appreciative to Professor Toni Schenk for being patient to academically revise this thesis and giving all the inspiration and critical guidance in the limited time he could spare for me.

I would like to acknowledge all my collegues at ifp wo have helped me in countless ways, in particular the members of the Photogrammetry and Remote Sensing Group, especially Heiner Hild and Christian Staetter for being very close companions. I wish them a lot of success in their careers. I am highly grateful to Norbert Haala who started me in this direction of research, and I benefitted from his expertise in this topic. I would like to thank Monika Sester for her encouragement and many discussion that we had on different aspects of Geomatics. I wish to thank Michael Kiefner, Karl-Heinrich Anders, and Dirk Stallmann who were always available for discussion on the programming aspect of this research study. My grateful thanks to Markus Englich for supplying me with computer hardware and software, Werner Schneider for providing the data set for testing, and Dorothee Klink for kindly editing this thesis in a significantly short time. Many thanks also go to Martina Kroma for various types of administrative support during these four years of study and research.

I wish to thank the National Cartography Center of Iran (NCC), in particular to Mr. Eng. Ahmad Shafaat former NCC Director, and Mr. Eng. Ali-Akbar Asgarian former Head of International Education Office of NCC who initiated and arranged the financial support for the first year of this research study.

To all my family and friends, in particular my brother Iraj Ameri, I acknowledge your support and encouragement directly and indirectly for the last four years during which I was on this journey of discovery.

Last but not least, I wish to express my unfathomable gratitude to my wife Matin Bayani for her understanding and her enduring patience during this very trying period. I could have never completed this thesis without her greatest support. I am in great debt to her in taking a lot of time from her to concentrate on this study during these years.

CurriculumVitae

Name	Babak Ameri Shahrabi
Jan. 24, 1966	Born in Tehran, Iran
1972 - 1977	Primary School, Tehran, Iran
1977 - 1980	Secondary School, Tehran, Iran
1980 - 1984	High School, Tehran, Iran
1984 - 1988	Department of Surveying, University of K. N. Toosi, Tehran, Iran Degree: B.Sc. in Surveying, Geodesy and Photogrammetry
1989 - 1991	Surveying Office, Food Industries Group-Guard Corps, Tehran Iran
1991 - 1992	Farazamin Consulting Engineers Co, Tehran Iran
1992 - 1993	Department of Geoinformatics, ITC, Enschede, the Netherlands Degree: Postgraduate Diploma in Production Photogrammetry (with Distinction) Degree: M.Sc. in Integrated Map and Geoinformation Production (with Distinction)
1992 - 1996	National Cartographic Center (NCC), Tehran, Iran
1996 - 2000	Institute of Photogrammetry, University of Stuttgart, Stuttgart, Germany Degree: Ph.D. in Photogrammetry and Computer Vision