D I S S E R T A T I O N

Automated Generation and Updating of Digital City Models using High-Resolution Line Scanning Systems

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Prof. Dr. Josef Jansa
E122
Institut für Photogrammetrie und Fernerkundung

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Georg Vozikis
9826142
1060 Wien, Bürgerspitalg. 2/8

Wien, am 23.05.2005 Unterschrift

Zusammenfassung

Digitale Stadtmodelle zählen zu den begehrtesten Produkten der Photogrammetrie und Fernerkundung. Die Erstellung und Aktualisierung dieser Modelle gehört zu den zeitaufwendigsten und teuersten Aufgaben, die in der Regel manuell durchgeführt werden. Ziel dieser Arbeit ist es eine automatisierte Methode zu entwickeln, die möglichst selbständig digitale Stadtmodelle erzeugt und aktualisiert. Als Grundlage sollen panchromatische Stereo-Bilder von hochauflösenden flugzeug- und satelliten-getragenen Zeilenscannern dienen.

Die vorgeschlagene Vorgehensweise kann in 3 wesentliche Schritte eingeteilt werden: Datenvorbearbeitung, Gebäudeextraktion (entspricht der Erstellung des Stadtmodells) und Feststellung von Veränderungen sowie Aktualisierung alter Datenbestände. Der erste Schritt, der die Akquisition und Orientierung der Bilddaten, sowie die Erstellung von Oberflächenmodellen und den daraus abgeleiteten normalisierten Geländemodellen beinhaltet, ist nicht Teil dieser Arbeit. Bei der Gebäudeextraktion werden vorerst Regionen gesucht in denen potentielle Gebäudekandidaten existieren könnten.

Dies geschieht durch eine Grauwertanalyse im Bild und einer Höheneinschränkung im normalisierten Geländemodell. In Folge werden die geometrischen Eigenschaften der gefundenen Gebäude mit Hilfe eines adaptiven Region Growings und einer darauffolgenden Hough-Transformation extrahiert. Die errechneten Gebäudeecken werden in einer Datenbank abgelegt und können zur Auffindung von Veränderungen verwendet werden.

Da diese Methode auf Gebäude, die eine Mindestgrösse von ca. 25 Pixeln haben eingeschränkt ist, wurden weiters zwei alternative Ansätze untersucht: Gebäude-extraktion mittels Bildmatchings oder mittels Texturanalyse.

Die Qualitätsanalyse der Ergebnisse zeigt, dass die untersuchte Methode des adaptiven Region Growings mit der darauffolgenden Hough-Transformation, sehr erfolgreich ist. Das Bildmatching liefert in kleinmasstäbigem Bildmaterial gute Resultate. Die Methodik basierend auf der vorgeschlagenen Texturanalyse ist hingegen nicht empfehlenswert, um einzelne Gebäude zu detektieren und extrahieren.

Acknowledgements

First of all I want to thank Josef Jansa, my mentor, who never got tired reading the 'TABU field-guide' and was always there to provide professional input, not only from the technical point of view, but also regarding formal matters. Andreas Georgopoulos who agreed to evaluate the thesis should also be mentioned. Moreover I want to thank Georg Gartner who was willing to examine me during the Rigorosum.

A big thanks also goes to the Austrian Academy of Sciences (ÖAW), which granted me a DOC scholarship, through which it was possible for me to finish my work in a time frame of 2 years.

Table of Contents

1 Introduction 1
1.a Problem Statement 2
1.b Motivation and Outline 5
1.c TABU – Tool for Automated Building(s) Updating 7
2 State of the Art 8
2.a Segmentation 8
2.a.i Segmentation of Image Data 8
2.a.ii Segmentation of DSMs 13
2.b Automated Building Extraction 15
2.b.i Automated Building Extraction from Image Data 16
2.b.ii Automated Building Extraction from DSMs 18
2.c Semi-Automated Mapping and Building Reconstruction 19
2.d Change Detection and Updating 20
3 Sensors and Data 24
3.a Spaceborne Digital Acquisition Systems 24
3.a.i QUICKBIRD (DigitalGlobe) 26
3.a.ii IKONOS (SpaceImaging) 26
3.a.iii Orbview3 (OrbImage) 27
3.a.iv SPOT5 (SPOT Image) 27
3.b Airborne Digital Line Scanners 27
3.b.i HRSC-AX (DLR) 30
3.b.ii ADS40 (LH-Systems) 30
3.c Used Data 30
4 Proposed Workflow 35
4.a Pre-Processing 36
4.a.i Orientation 36
4.a.ii DSM Extraction 37
4.a.ii.A General Remarks on Image Matching 38
4.a.ii.B Area-Based Matching 39
4.a.ii.C Feature-Based Matching 39
4.a.ii.D Structural Matching 40
4.a.ii.E Matching Problems 40
4.b DSM Normalization 40
4.b.i nDSM Generation – State of the Art 41
4.b.ii nDSM Generation in TABU 44

4.b.iii nDSM Generation – Example 47
4.c Building Detection 48
4.c.i Building Detection - Examples 50
4.d Building Extraction 52
4.d.i Adaptive Region Growing 53
4.d.ii The Hough Transformation 55
4.d.iii Edge Detection (Hough Back-Transformation) 60
4.d.iv Corner Determination 63
4.d.v Strengths of the Hough Transformation, Examples 66
4.d.v.A Noisy Data 66
4.d.v.B Level of Detail 67
4.d.v.C Forcing Angles 69
4.d.v.D Bridging Gaps 72
4.e Change Detection and Updating 73
5 Alternative Approaches 76
5.a Building Extraction by Image Matching 77
5.b Building Extraction with Texture Analysis 82
6 Results and Quality Assessment 87
6.a Quantitative Assessment of Building Extraction 89
6.b Qualitative Assessment of Building Extraction 90
6.c Problems 92
6.c.i Heterogeneous Roofs 93
6.c.ii Shadow Effects 94
6.c.iii Complex Buildings 94
6.c.iv DSM Inaccuracies 95
6.c.v Adjacent Buildings 96
7 Conclusions 99
8 Acronyms 103
9 References 105

1 Introduction

Digital City Models (DCMs) have become one of the most important and attractive products of photogrammetry and remote sensing. The spectrum of application areas dealing with such models is huge: environmental planning and monitoring, location based services, navigation, virtual reality, cadastre, 3D map updating etc. Moreover they can be used for the derivation of propagation patterns (e.g. noise pollution, radio waves for telecommunication, flood waters and many more).

When looking at photogrammetry and remote sensing, one can observe that during the last decade these sciences have experienced a big technological advance. On the one hand, because the geometric resolution of satellite imagery has increased tremendously (footprints of one metre or less are current state); on the other hand, photogrammetry has become more and more 'digital' and new techniques are being developed to accelerate the workflow. Due to this progress the science of satellite remote sensing starts dealing with application areas, that until now where classical photogrammetric tasks, such as large scale mapping or city model generation.

The change from analogue/analytic to digital photogrammetry enabled the automation of many traditional (time-consuming) photogrammetric tasks, such as

- measurement of fiducials (automation of inner orientation) [Schickler and Poth (1996)],
- automatic measurement of tie-points (automation of relative orientation) [Heipke (1997)],
- automatic measurement of control points (automation of absolute orientation) [Schickler (1992)],
- and automatic DSM (Digital Surface Model) extraction [Hannah (1988)].

These automated procedures have become standard and are included in most of the commercial digital photogrammetric workstations.

1.a Problem Statement

Regarding the automation of feature extraction (buildings, individual trees, roads etc.), it seems that no flawless procedure has been found yet. There exist many difficulties at various stages. Granzow (2001) concludes that automated feature extraction cannot be expected to serve as a final product. It should be used to facilitate image analysis and interpretation as a supplemental technology.

The term 'automation' is quite often misleading. In order to categorize the degree of automation one should find an answer to the following questions. How much user interactivity is required when working with a (semi-) automatic/automated system? What is the potential of such a system?

According to Guelch (2000) four levels of automation can be achieved:

1. Interactive systems: No Automation, all measurements are performed manually.
2. Semi-automatic systems: Automatic modules are integrated in a more or less interactive workflow. Approximate values are required for the automatic modules.
3. Automated systems: Main tasks are performed automatically. User interaction is focused on project setup, post-editing, correction of erroneous results and on the definition of the parameters that will be used in the automatic process.
4. Autonomous systems: Fully automatic *"press-button-and-forget"* systems. Here the system calculates the parameters for the automatic processing autonomously. Such systems do not yet exist for photogrammetric applications. [Rottensteiner (2001)].

Since the approach for creating an automated system highly depends on the spatial resolution of the input data, let us define what exactly we mean when speaking of different resolutions. The term "high-resolution" has always been associated with the best known sensor at that time. Thus ten years ago, high-resolution satellites provided imagery with a GSD (Ground Sample Distance)[1] of ca. 10m.

[1] The term Ground Sampling Distance (GSD) is used here for defining the spatial resolution of an image. By "spatial resolution" one understands the capability of an imaging system to separate two neighbouring objects of a certain spatial extent. The resolution is defined in terms of lines or line pairs (i.e. a sequence of a dark and bright bar) per unit distance, e.g.

In 1999 when the IKONOS satellite was sent into orbit providing a ground resolution of 1m, things started to become confusing. Many authors add adjectives like "very high" or "ultra high" to make a clear differentiation. Ehlers (2002) proposes the following taxonomy:

- Ultra High: < 1m GSD
- Very High: 1-4m GSD
- High: 4-10m GSD
- Medium: 10-50m GSD
- Low: 50-250m GSD
- Very Low: >250m GSD

In order not to irritate the reader with all these differentiations, in the following a GSD of 2.5m should be categorized as "high-resolution". For a GSD up to 50m the term "medium resolution" should be used.

Methologies vary dependent on whether medium resolution images or high resolution images are utilized for urban change detection. Two main differences may be mentioned. While medium resolution images are not appropriate for recognizing individual buildings, they may provide sufficient information for the segmentation into urban and non-urban regions. If a series of multi-temporal images is available, the dynamics of urban development can be derived. Common approaches for the segmentation procedure are:

1. Classification for multispectral data, e.g. [Kraus (1990)],
2. Texture analysis for panchromatic data, e.g. [Haralick and Shapiro (1985)],
3. Spectral mixture analysis for multispectral data, e.g. [Lu and Weng (2004)],
4. or a combination of the listed methods.

High-resolution data are best suited for the extraction of the geometric properties of single buildings. If stereo-imagery or other additional height information is provided, building heights may be derived and subsequently (3D) DCMs may be created.

---

lines per metre [L/m]. The sampling theory explains the dependence between resolution and sampling. One must not forget, that the Instaneous Field of View (IFoV) also influences the spatial resolution, which is defined as the object area covered by a sensor element (often referred to as "footprint") from which recorded reflectance value is derived. In the ideal case GSD and IFoV are identical, which should be assumed in the following.

There is a great variety of techniques for building extraction, ranging from sophisticated edge detection techniques and fuzzy interpretation to the use of Bayesian networks. An overview is given in section *2.State of the Art*.

Besides image acquisition airborne laser scanning has become an important technique for 3D data capture. Also known as laser altimetry, LiDAR (Light Detection and Ranging) or LRF (Laser Range Finder) nowadays it is used in diverse (also non-cartographic) fields [Axelsson (1999)].

In the last years laser scanning has reached maturity due to the enhanced performance of GPS and IMU (Global Positioning System and Inertial Measurement Unit) for capturing position and orientation data [Ehlers et al. (2003)]. Hence it has become the leading method for creating high density DSMs.

When creating DCMs a first great challenge is the initial detection of the presence or absence of a building in the image region. Locating buildings reliably and completely in images within urban areas that contain hundreds of objects such as parking lots, fields, road networks and vehicles, just by utilizing their spectral properties, is not possible [Jaynes et al. (2003)]. Additionally, trees, power lines and other buildings often occlude building roof tops. Moreover, there exists a great variety of shapes of building roof tops.

Whereas most of the approaches aim at segmenting the area under investigation into building and non-building regions by analyzing elevation (either derived from LiDAR data or by image matching), NDVI (Normalized Difference Vegetation Index) or textural measures [Kim and Muller (2002), Haala and Brenner (1999), McIntosh and Krupnik (2002), Vosselman et al. (2004), Rottensteiner and Briese (2002)], others employ so-called Support Vector Machines (SVM) [Bellman and Shortis (2004)] or apply graph analysis on extracted edges in the imagery [Jaynes et al. (2003)].

Newer, hybrid methods concentrate on processing laser scanner data and optical imagery simultaneously, since a combination of both technologies provides the best solution regarding automated building extraction [Maas (2004)]. The FLACON system for example, acquires elevation data and multispectral imagery concurrently [Lohr (2003)]. Its laser scanner, which is position to capture first and last pulse data, operates with a fibre glass array producing a parallel acquisition pattern. The imaging sensor captures four bands in the visible and near infrared spectrum.

Many ideas and implementations for highly accurate extraction of individual buildings from laser scanner and/or image data have been published.
Nevertheless, it does not seem that there has been found the "perfect" automated solution yet, especially when dealing with (high-resolution) images. Either the correctness, reliability and consistency of the results is not satisfying enough or the degree of automation is too low.

1.b Motivation and Outline

Manual updating of old maps or DCMs has always been one of the most time consuming and expensive tasks. Many concepts were developed to support the operators in their work, but still human interaction is very high.
This PhD-thesis focuses on the automation and acceleration of the production of DCMs. Furthermore, it is tried to increase their quality and to keep them up-to-date.

As input neither laser scanner data, nor multispectral images are used. The only data source are high-resolution panchromatic stereo-images, either taken from high-resolution satellites (e.g. Quickbird, IKONOS, Orbview, SPOT5) or by airborne line scanning systems (e.g. HRSC, ADS40). Due to this limitation our gaol becomes even more challenging.

The idea of using only panchromatic imagery arose from the fact that the panchromatic bands of remote sensing satellites are captured in high spatial resolution, whereas the multispectral data is usually 2 to 5 times coarser (see Table 3-1). The same holds for airborne digital acquisition systems, both step-and-stare scanners (frame cameras, e.g. DMC [Hinz and Doerstel (2001)], Ultracam [Leberl et al. (2003)]) and pushbroom sensors (line scanning systems, e.g. ADS40 [Fricker (2001)], HRSC [Neukum (2001)]).

The objective is the development of an automated prototype system where user interactions are minimized and the reliability of the outcome is maximized. Through the automated creation of a DCM which represents the current (new) state, a comparison and thus change detection can be performed with the old state that can be represented through an old DCM, digital cadastre or other vector database. The detected changes can be directly registered in the database.

Urban change detection is very important, not only for a better understanding of complex territorial problems, but also to provide a complete basis for the approaches

to urban and regional spatial planning and to help city managers and decision-makers in defining local policies [Kemper et al. (2004)].

The proposed approach can be divided into five major steps:

1. Pre-processing (data acquisition, orientation, matching)
2. DSM normalisation
3. Detection of potential building candidates
4. Extraction of geometric properties of found buildings
5. Change Detection and Updating.

The orientation and DSM extraction of spaceborne stereo images is carried out with the LPS (Leica Photogrammetry Suite) software version 8.7 by Leica Geosystems. The airborne data tested in this PhD research was already pre-processed by company ISTAR or DLR (German Aerospace Centre).

The normalization of the DSM is an important step for further processing. The so-called nDSM (normalised DSM) is the difference between the DSM and the DTM (Digital Terrain Model) and thus provides a height indication for all objects like buildings protruding from the terrain. The proposed algorithm is simple, efficient and automated to a very high degree (see *4b.DSM Normalisation*).

The finding of potential building candidates is carried out by analysing textural image features as well as the elevation of the nDSM.
Once the location of a building is detected, the actual building extraction (derivation of the building's corners), is done by employing an adaptive, iterative region growing process in combination with sequence of Hough transformations. The use of the Hough transformation proves to be very favourable since gross errors from the region growing process can be corrected. Furthermore, the proposed implementation of the Hough transformation allows various generalisation possibilities and geometric constraints (see *4c.Building Detection*).

For images where the building sizes are below a dozen pixels this procedure fails; thus a solution is suggested based on pattern recognition and textural analysis (see *4d.Change Detection and Updating*).

Finally, the change detection and updating step compares the newly created DCM with the old state (e.g. old database). Multiple possibilities for describing and validating the found changes are investigated (see *4e.DSM Normalisation*).

In our DCMs each building is assigned one single height value and hence is considered as a building with a flat roof. The used, matched DSMs are too coarse to provide information for modelling detailed roof structures.
The accuracy of the derived building corners depends on the spatial image resolution, quality of triangulation and on radiometric aspects.

1.c TABU – Tool for Automated Building(s) Updating

The prototype software TABU (Tool for Automated Building(s) Updating) that has been developed as working and testing tool is written in IDL (Interactive Data Language). TABU consists of the following main modules:

- standard image processing: focal filtering, histogram filters, contrast enhancements, adaptive filters, morphological filters, edge operations
- nDSM creation
- adaptive region growing
- vectorization
- module for edge and corner extraction
- pattern recognition, template matching, texture analysis
- Foerstner operator
- module for transforming various image and vector formats into each other
- tools for 2D and 3D visualization

2 State of the Art

This chapter gives an insight into state of the art techniques for deriving DCMs, as well as change detection and updating procedures. Most of the methods can be divided into two major steps, namely the detection of a building in the dataset and the extraction of its geometric properties.

2.a Segmentation

Segmenting an image is the process of labelling the pixels so that pixels whose measurements are of the same class are given the same label [Hoover et al. (1996)]. It is one of most the difficult tasks in image processing and its accuracy determines the eventual success or failure of computerized analysis procedures [Gonzalez and Woods (2002)]. The most common techniques for separating the image area under investigation into building or non-building parts are based on the analysis of multispectral information or the NDVI and DSM. The NDVI ratios biomass response in the near infrared against biomass absorption in the visible red part of the spectrum [Morain and Baros (1996)].

Generally speaking the DSM could be used for distinguishing between terrain and above-terrain points and the NDVI provides information for segmenting the above-terrain points into buildings and vegetation. When dealing with laser scanner data the DSM investigations are a little bit different.

2.a.i Segmentation of Image Data

Hoffmann et al. (2001) show that the segmentation and classification problem is not only an interesting research topic, but also very important in practice. A good example is the MOLAND (Monitoring Land cover/use Dynamics) project, which is carried out by the Institute for Environment and Sustainability of the JRC (Joint Research Centre) of

the European Commission [https://moland.jrc.it, last accessed on 25 February 2005]. The three specific aims are [Kemper et al. (2004)]:

1. Produce quantitative information in the evolution of land use and transportation networks, from 1950 onwards.
2. Develop methods for performing a harmonised analysis of historical trends, including socio economic aspects, impact of legislation, landscape fragmentation etc.
3. Develop models for the harmonized simulation of future European-wide scenarios, at local and regional scales.

The goal is to provide a spatial planning tool that can be used for assessing, monitoring and modelling urban and regional development. In the MOLAND project many techniques have been tested (from simple multispectral classification up to fuzzy classification approaches) and partly excellent results are achieved.
Unfortunately, these investigations are focused on rather small scales and deal with the classifying of urban regions. Extraction of individual buildings is not of interest.

If multispectral imagery is available the classification approach is the most convenient way for detecting building areas or urban regions. Knudsen and Olsen (2003) apply an unsupervised spectral classification method that controls a subsequent Mahalanobis classification step [ERDAS (2003)]. Due to good geometric resolution of the used RGB images, changes can be traced and even new buildings detected.

Orun (2004) evaluates non-linear classification techniques by the unification of appropriate texture analysis methods and a learning Bayesian classifier [ERDAS (2003)].

In many cases stereo imagery is available. The derived DSMs and resulting nDSMs can also be used as an important source for performing the segmentation, although most of the times they are very coarse if compared to laser scanner DSMs. Most of the times, simple thresholding is applied on the nDSM for separating terrain from above-terrain points. The big problem remaining is the segmentation of the above-terrain points into building and vegetation points. If multispectral imagery is disposable, the most popular method for separating buildings from vegetation is to use the NDVI [Morain and Baros (1996)], which clearly states whether any vegetation exists or not.

Kim and Muller (2002) for example choose this approach for segmentation. Everything that exceeds a certain threshold in the nDSM will be included, and vegetation will be

afterwards excluded by interpreting the NDVI. The same procedure is followed by Sohn and Dowman (2002) and Trinder et al. (2002).
Olsen (2004) also makes use of the nDSM and NDVI in order to validate training areas for a further classification step, but in the clustering process some classes are subdivided into more unique subclasses since they are spectrally diverse. Subclasses, which are spectrally more uniform than the base building class, are used to perform the actual classification of the entire image (see also Kressler and Steinnocher (1996)).

In order to get a rough impression where building candidates are, Niederoest (2000) performs a segmentation by multiplying the nDSM with the so-called 'Degree of Artificiality' (DoA), which is a function of the red and the green band. In the outcome bright pixels are classified as man-made objects (high DoA).

Another approach to segmentation focuses on texture analysis, specially when no multispectral data is available. Although no formal definition of texture exists, intuitively descriptors provide measures of properties such as smoothness, coarseness and regularity [Gonzalez and Woods (2002)].
The three principal approaches used in image processing to describe texture are:

1. statistical: characterizes the texture as smooth, coarse, grainy etc.
2. structural: these techniques deal with the arrangement of image primitives, such as the description of texture based on regularly spaces parallel lines,
3. spectral: based on properties of the Fourier spectrum and are primarily used to detect global periodicity in an image by identifying high-energy, narrow peaks in the spectrum.

The most common approach is the first one. Known techniques are based on:

- grey level co-occurrence statistics [Haralick (1979)],
- grey level run length statistics [Chu et al. (1990)],
- texton gradients [Julesz (1986)],
- Gabor filters [Jain and Karu (1996)] and
- random field models [Wu and Doerschuk (1994)].

Co-occurrence measures are widely used for texture analysis. Zhang (2001) shows that a multispectral approach with integrated textural information can increase the accuracy. His goal is to detect trees (vegetation) and separate them from other man-

made objects (especially buildings) in high-resolution images. Therefore he makes use of this texture integrated classification approach. The textural features he uses are Contrast, Energy, Entropy, Homogeneity and the "Conditional Variance". The problem with the typical variance measure is that in the operation window high values not only indicate heterogeneous regions (vegetation), but also noise (edges of buildings, shadow-light changes etc.). To avoid this shortcoming Zhang (2001) implements a unidirectional variance detector (Conditional Variance). "...the pixel variance on each side of the central pixel is first calculated. The central pixel is then assessed according to the values of the unidirectional variance on each side. When the variance value on one of the four sides is less than a threshold, it means that there is a homogeneous area or line on this side. In this case, the central pixel should be located either in a homogeneous area or on a straight edge, but not inside a tree area..." [Zhang (2001)].
The result accuracy of a simple multispectral classification was 67%, and could be increased up to 96% by employing the texture integrated classification method.

Sindhuber (1998) also suggests the combination of a multispectral classification and texture andalysis in order to impreove classification results in the following way:

- a multispectral Maximum Likelihood classification, where for each pixel the classes with the highest and the second highest probability are stored
- and a textural analysis using the Foerstner operator.

This information is fused through a rule-based system.

Carbal et al. (2004) provide one more example for processing multiple data sources, namely multispectral and textural information. A first segmentation is carried out by analyzing the NDVI and for refinement a second segmentation is carried out by applying textural filters.

Similarly, Ehlers et al. (2003) suggest controlled region growing by a comparison of the obtained homogeneity measures of neighbouring primitives with a user defined threshold. Afterwards a classification through a fuzzy-logic approach is carried out in order to interpret the obtained segments.

The number of image measures (spectral, textural, morphological) is huge. Calculating all these measures for a given data set would increase the processing

time tremendously. Furthermore, many of these measures are highly correlated; hence a problem of separabilities might arise.

Carleer and Wolf (2004) investigate this problem by examining which features seem to be most suitable for the analysis to separate vegetation from non-vegetation. 4 bands of multispectral spaceborne imagery are used and 33 measures (10 spectral, 7 morphological and 16 textural) are tested. The goal is to reduce this number to 5 and at the same time achieve the same results (remove features with high correlation). The most suitable features are found by computing the Bhattacharyya distance B (Equation 1-1).

$$B = [\mu_1 - \mu_2]^T \left[\frac{\Sigma_1 + \Sigma_2}{2}\right]^{-1} [\mu_1 - \mu_2] + \frac{1}{2} \ln \frac{\left|\frac{1}{2}[\Sigma_1 + \Sigma_2]\right|}{\sqrt{|\Sigma_1||\Sigma_2|}} \qquad (1\text{-}1)$$

where μ is the mean vector and
Σ is the covariance matrix of a class.

The most appropriate set of independent measures reported in Carleer and Wolf (2004) is listed below (ratio of a band = the ration between the spectral band and the sum of all spectral bands):

- mean of NDVI,
- ratio of red band,
- ratio of near infrared band,
- angular second moment of panchromatic band.

This result shows clearly how valuable multispectral information is for proper image segmentation; especially for distinguishing between vegetation and non-vegetation. It also shows that in general the correlation between textural features is much higher than it is between spectral measures.

An alternative classification approach is proposed by Bellman and Shortis (2004). They introduce so-called Support Vector Machines (SVMs). The method is based on the principle of risk minimization [Vapnik (1995)] and has the attractive property that it minimizes the bound on the generalization error. The SVM is trained on a large sample of building and non-building regions, and is used afterwards to classify

previously unseen image patches. A success rate of 80% for large-sale imagery is reported.

Another promising but rarely used data source are hyperspectral airborne data (e.g. from the scanners AVRIS, DAIS, HYMAP, CASi). They offer excellent conditions for segmentation and classification since they provide high spatial, spectral and radiometric resolution. Unfortunately, these sensors are only limited used because of their enormous costs.

2.a.ii Segmentation of DSMs

Airborne laser scanning has become one of the most important acquisition techniques for surface reconstruction. By collecting a huge amount of points (3D locations) with a very high resolution and accuracy they provide an optimal input for building extraction algorithms. All of these algorithms need initial building information; it has to be approximately known where a building is located, in order to start the extraction procedure. Many times, this initial information is taken from (digitized) 2D ground plans [Peternell and Steiner (2003), Brenner and von Goesseln (2004)]. In case such initial data is not available a segmentation procedure has to be carried out to separate building points from non-building points. Here the segmentation approaches are rather different to those applied to aerial and spaceborne imagery. The major problem is the separation of building from vegetation points here too.

Very useful information for this distinction is the difference between first and last pulse of the laser scanner. Over buildings this difference is zero, whereas over vegetation it is not equal zero, because some of the laser points are reflected on the canopy and others manage to reach the ground.

Rottensteiner (2004) distinguishes between building points and other off-terrain points (e.g. vegetation) by applying morphological operators and texture analysis on the DSM, since we assume that building points behave in a more homogeneous way in the DSM than vegetation points do.

Similarly, Baillard and Maitre (1999) segment the whole scene into homogeneous and heterogeneous parts by carrying out a region growing and merging algorithm.

Haala and Brenner (1999) employ a segmentation technique based on homogeneous local surface normals. The normal direction is approximated using the derivatives of a local bivariate polynomial fit.

Maas (1999) suggests extracting texture measures from height co-occurrence matrices. These measures together with the information of first and last pulse and the heights of the nDSM are used for an unsupervised K-means classification[2]. Maas (1999) claims to achieve 90-95% correct results.

Matikannen et al. (2003) use a bottom-up region merging algorithm to create segments that are associated with texture measures. Afterwards a fuzzy classification is performed. Fuzzy classification methods take into account mixed pixels. A thorough discussion on fuzzy classification is given in ERDAS (2003). 90% correct classification results are reported.

Voegtle and Steinle (2003) make also use of a region merging algorithm and a subsequent fuzzy classification. Attributes for segments are gradients of segment borders, differences between first and last pulse, and shape and height texture measures. 93% positive results are stated.

Vosselman et al. (2004) too propose to take the difference between first and last pulse into consideration, as well as the DSM roughness and segment sizes. It is reported that 85% of building points and 78% of vegetation points are classified correctly.

In Roggero (2002) another interesting approach is discussed. The segmentation is carried out by combining region growing with a principal component analysis (PCA):

- Hierarchic region growing: points are merged if they are close to each other, not only in sense of Euclidian distance, but also in terms of reflectance or geometrical properties. Start point is a randomly chosen non-aggregated point. This is done until every point belongs to one region.
- Refinement: this procedure is again carried out up to four times for each segment, hence points in one segment are subdivided into multiple groups.
- The PCA is then used to analyse the structure of each data set in a segment.

[2] Simply speaking K-means clustering is an algorithm to classify or to group your objects based on attributes/features into K number of groups. K is a positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus the purpose of K-mean clustering is to classify the data.

- Allegedly this procedure is very robust and works well in noisy environments. (For further reading on the PCA see Kraus (1990) or Gonzalez and Woods (2002).)

Most of the above mentioned methods mainly aim at distinguishing between vegetation and building points. The separation between terrain and above-terrain points can be carried out by employing the normalized surface model and applying a simple elevation thresholding. This threshold value is usually based on empirical knowledge and depends on the building heights of the area under investigation. The threshold should be set not higher than the lowest building that has to be modelled.

An idea of how to automatically derive this threshold value is described in McIntosh and Krupnik (2002). The segmentation between above-ground (building/vegetation) and ground points is carried out by an empirically derived threshold T:

$$T = h + 0.75\sigma \tag{1-2}$$

where h is the mean and
σ the standard deviation over all laser points.

It is reported that equation (1-2) holds for flat urban regions with one to two storeys buildings.

For further reading on segmentation algorithms see also Hoover at al. (1996).

2.b Automated Building Extraction

Once the locations of buildings have been found, either by automated segmentation or manual digitization, the building extraction process can be started. Here too, the approaches differ depending on whether image data or laser scanner data have to be processed, mainly due to the differences of the nature of the data.

On the one hand, building extraction in high-resolution imagery deals with deriving the building outlines in order to obtain the building corners. On the other hand, when working with highly dense point clouds from LiDAR systems, most approaches focus on extracting roof planes and intersecting them. These models can become very detailed and in combination with accurate ground plans, even roof overhangs can be modelled.

One major problem when using laser scanner data is the distinction between neighbouring buildings if their orientation, their height and their shape, their roofs, are identical. Individual buildings cannot be separated from each other just by analysing DSM data. Neighbouring roofs may be distinguishable in images if the borderline is radiometrically visible (i.e. if the colour or texture is different). Hence many researchers apply fusion approaches to laser scanner data and imagery in order make use of complementary information [Rottensteiner (2003), Ehlers et al. (2003), Sohn and Dowman (2002), Trinder et al. (2002)].

2.b.i Automated Building Extraction from Image Data

An algorithm applied to aerial stereo imagery (scale 1:40000) is proposed by Avrahami et al. (2004). It consists of five steps:

1. pre-processing,
2. left image operations,
3. height operations,
4. right image operations and
5. mapping in object space.

The user has to manually determine a point on a building roof in the left image. The 2D contour around the roof is extracted using a region growing method. Now, for each point on the contour the homologue point is searched in the right image on a specific area limited by geometric conditions: the epipolar line and the possible maximum and minimum height of the building. The exact position of these homologue points in the right image is determined by means of radiometric difference modelling. As the building elevation is now known, the building contour is transferred to 3D object space.
The average building mapping (ABM) rate is 76% and described by Avrahami et al. (2004) as follows (Equation 1-3).

$$ABM = \frac{BSM + k \cdot BPM}{BSM + BPM + BNM} \times 100 \qquad (1\text{-}3)$$

where BSM number of buildings successfully mapped,
BPM number of buildings partially mapped,
BNM number of buildings not mapped and
k is a weight for evaluating success in the BPM class and set to 0.5.

The planimetric accuracy of the extracted edges is 0.5m and the height accuracy amounts to approximately 1m. It has to be mentioned that in this case study the roofs are very homogeneous, no shadow effects or occlusions exist and the buildings are isolated.

Lhomme et al. (2004) discuss an approach for automatic building extraction from images that consists of a zonal grey value analysis and a per pixel classification.

Jaynes et al. (2003) present a model-based approach in order to automatically detect and reconstruct buildings from aerial stereo imagery and a matched DSM. To obtain the building outlines the following process was pursued:

- Low-level feature extraction: obtain straight lines and corners by applying an edge detector.
- Collated feature detection: corners and line segments are grouped into chains of features that represent a polygonal boundary. These chains are stored in a so-called feature relation graph.
- Hypothesis arbitration: obtain rooftops by parsing the feature relation graphs for closed chains that are both independent (have no edges in common) and of maximal certainty. (The chain of collected features inherits an accumulated certainty measure from all the nodes along its length. Nodes have a certainty measure related to the confidence of the low-level feature extraction routines, and edges are weighted with the certainty that a grouping represents.)

Once the rooftops are determined, each is compared to the entries of a database of parameterized surface models that represent different building shape classes. Then, through an iterative process the precise position of the roof is determined. The idea of matching a database of shape models against each segment is utilized.
It has to be mentioned that this method was applied to big (factory) buildings, therefore the matched DSM was good enough for this job.

Taillandier and Deriche (2004) apply a generic Bayesian network to reconstruct buildings from aerial stereopairs. The spatial resolution of the imagery is 25cm. The focusing zone, i.e. the region where a building is located, is selected manually. This individual building reconstruction algorithm consists of three steps:

1. Primitive detection (look at locations of vertical or nearly vertical walls).
2. Hypothesis extraction.

3. Choice if the best representation and geometric refinement.

It is stated that 75% of the reconstructions are "...acceptable, meaning they perfectly fit the reality or the given caricature is an acceptable generalization of reality." [Taillandier and Deriche (2004)].

Sohn and Dowman (2002) combine IKONOS imagery with a rather coarse laser scanner DSM (3m resolution) for extracting roof outlines. The buildings are reconstructed in form of polyhedral shapes. After a hierarchical segmentation of the LiDAR data in combination with the NDVI of the imagery, intensity line cues are extracted from the imagery and verified through the LiDAR data. Virtual line cues are generated for generating the polyhedral shapes.

2.b.ii Automated Building Extraction from DSMs

Throughout the world approaches have been presented for automatic building extraction from laser scanner data, among them one by Rottensteiner (2003). He performs three major steps. The first two steps deal with the segmentation of the point cloud into roof and non-roof points. The third step is dedicated to the geometrical reconstruction of the buildings:

- The geometric parameters of planar patches are derived together with the RMS (Root Mean Square error) from all points inside a homogeneous region.
- Patches with the best fit (with respect to a least squares solution) are considered as seed regions for the following region growing process (Neighbouring pixels are iteratively accumulated if their distance from the original plane is below a certain threshold.)
- Computation of improved plane parameters for final model.

A different method for automated building reconstruction is described in Peternell and Steiner (2003). Assuming that a dense point cloud is given, they try to determine all planar regions of a building. Firstly, local regression planes for all points are computed and their analytical parameters are transformed into a parameter space. Roof planes can be detected by finding clusters in the parameter space. The roof shapes are derived by intersecting and analysing the roof planes. Building boundaries are taken from ground planes of a GIS.

Both above mentioned approaches need a dense and highly accurate point cloud in order to be able to derive and analyse the roof faces of buildings. They aim at rather detailed reconstructions.

2.c Semi-Automated Mapping and Building Reconstruction

Automatic city model generation is a hot research field and is far from being mature. No matter whether laser scanner point clouds or images are processed, the reliability of the results is either too low or the success rate is not satisfying for the requirements of practice.

We face two situations:

- In practical applications we want 100% reliable results that are complete and consistent. Automation is of secondary importance.
- Research approaches try to maximize the automation, even if the results are not correct to their full extents.

The golden mean of manual, interactive procedures and immature automated building reconstruction techniques are semi-automated mapping methods. The user is still responsible for every crucial step in the process, but is supported by automated procedures in order to reduce working time [Guelch (2000)].

Semi-automated mapping and building reconstruction procedures can be subdivided into two groups:

1. Systems based on interactive (stereo-) measurements of a structured point cloud, from which the building topology is derived automatically. The result is many times not unambiguous.
2. Systems based on the principles of Constructive Solid Geometry (CSG): buildings are reconstructed by a combination of parametric primitives by Boolean set operators.

Rottensteiner and Schulze (2003) propose a semi-automated building extraction technique from aerial stereo imagery using adaptable primitives. Buildings are reconstructed by a combination of parametric primitives that are approximately positioned in the images by a human operator. The user selects a primitive from a library of primitives with standard roof shapes and determines its approximate location, orientation and shape. Optionally, an automatic fine measurement can be performed where the primitive's edges are matched with edges extracted from the images . More complex buildings are composed of simple primitives which are combined by automatic intersection.

The Cyber City Modeler software [Ulm (2003)] is based on a different approach. This semi-automated method utilizes a topology generator that fits planar structures of the building roof to structured point clouds that are manually measured by a human operator. The structured point cloud describes the building roof. Before the automatic process commences, the captured points have to be classified manually into boundary points and interior points. Then the modeler creates roof faces and their accurate topology autonomously.

In this case a roof type library is not required. The disadvantage is that the topology derived from the so-called structured point cloud may not be unambiguous. Although roof modeling of the previously described method is restricted to the set of primitives stored in a library, the restriction is marginal as the final building may be composed by intersecting individual primitives. The latter method, in contrast, needs more operator interaction already during the measurement and also for post-editing after the automatic process.

Brenner and von Goesseln (2004) combine both approaches for modelling the inner city of Hannover. As input both aerial laser scanner data and aerial imageres were used. The program ATOP is used for building modelling from the aerial laser scanner data. Additionally, 2D ground plans of buildings, which provide strong information on building extents, are integrated.
In a first step, a fully automatic process generates an initial reconstruction for all buildings. The ATOP system basically relies on the decomposition of ground plan polygons into 2D primitives. Each 2D primitive is the footprint of a corresponding 3D primitive. The parameters of the roof (namely roof type defined in a fixed library), height of the building and roof slope are determined using a segmentation and least squares adjustment.
For relatively simple buildings, this reconstruction already reflects the 3D structure properly. For more complex buildings, rework is required, which is done by using a semi-automatic modeller. In this case the commercial software inJECT is used. This program uses semi-automatic measurements in several mono images in order to derive roof structures. As with the ATOP approach, the buildings are modelled by arranging primitives.

2.d Change Detection and Updating

Methods applied for change detection mainly depend on the fact whether "equal" or "unequal" data sources are compared. On the one hand, multitemporal images or two databases with entries from different dates can be compared to each other in

order to determine changes. On the other hand, an old database can be compared to a recently taken image.

If high-resolution imagery is available, where individual buildings can be extracted, the change detection procedure will be different to that, if only medium scale imagery could be employed. The procedure for change detection in the first mentioned case may be carried out in a rather simple way, by just comparing the newly extracted buildings with the old state of the buildings. For the second case, a great variety of change detection techniques using remote sensing data exist, such as differencing, image regression, image rationing, vegetation index differencing, principal component analysis, post-classification subtraction, tasselled cap transformation and neuronal networks [Kaufmann and Seto (2001), Singh (1989), Dai and Khorram (1999)].

Change detection has been studied using several approaches. In Jensen (1997) a detailed description of state of the art change detection problems and applications can be found. Generally speaking, three approaches are common:

1. Detection of a single class of objects using a comparison algorithm. This is appropriate if invariant measures exist for the class of objects to be detected. E.g. The DSM could be computed from the new images and compared to the heights stored in an old 3D database in order to detect new building constructions.
2. Changes for an entire scene are processed simultaneously. Here an efficient classification algorithm is required for almost every class of the scene.
3. Compare images using a given similarity measure without any information on objects belonging to the scene before and after this comparison.

Jung (2004) makes use of multitemporal aerial stereopairs and tries to detect building changes. The image locations which are likely to have changed will be submitted to a human operator who should check the proposed change and validate it in order to update a geographic database.
The algorithm consists of two steps. In the so-called focusing phase the whole image is segmented into unchanged areas and possibly changed areas by comparing the DSMs of two different epochs. In the second step the regions of interest are divided into buildings and non-buildings. This is done by a combination of several decision trees induced from training data. In the described case study it is stated that an operator to whom the final set of locations was submitted, omits less than 10% of true changes.

The ATOMI project [Eidenbenz et al. (2000)] is used to update vector data of building roof outlines and increase their accuracy by applying image analysis techniques on aerial true-colour images (scale of analogue images is 1:15800). For finding new buildings Niederoest (2000) tries to localize extracted DSM blobs or apply a multispectral classification. In order to extract DSM blobs that really correspond to buildings we have to assume that there exists no extensive and high vegetation and that the DSM resolution and quality is good enough to be able to separate buildings from each other or from trees. He also makes use of the DoA measure for separating man-made objects from the rest. It is reported that 89% of all already existing buildings were reconstructed (over an area with a rather low building density) and also some new buildings were found, but nothing is said about the geometric accuracy of the results.

Olsen (2004) too uses aerial RGB imagery (scale 1:25000) for updating a 3D vector map. Additional height information is taken from a 1m resolution laser scanner DSM. After a very thorough classification step where buildings have been assigned a single class, a change detection map is computed. This map is the result of a pixel-by-pixel comparison between a raster version of an existing map database and the classification result. To reduce noise in form of single pixels in this change detection map a morphological opening is performed.
Approximately 50% of the factual changes in the test area were detected by the proposed algorithm.

Matikanen et al. (2004) use similar input data (laser scanner data, aerial images and map data) and follow a likewise procedure. They state that 88% of all buildings and 98% of all buildings larger than 200m^2 were detected correctly.

Information fusion concepts are taken into consideration in Samadzadegan and Abbaspour (2004) for revising urban geospatial databases using high resolution satellite images. The approach takes advantage of feature level fusion of descriptive information and should be able to handle data from various sources. The methodology is briefly described below. The first input category consists of a DSM, textural information, and spectral information, all derived from the images. The second input category is vector data of a 3D digital map.

- Object identification: an object oriented approach for data structuring is suggested.
  - Definition of objects (descriptive parameters: height, shape, area, roughness etc.)

- generate so-called linguistic variables, membership functions for the fuzzy rules for the descriptive parameters.

- Object extraction:
  - preliminary extraction of all objects of interest (3D region extraction)
  - modification and verification of these objects based on all available information (structural, textural and spectral)

- Change detection / Map revision: comparison of objects detected on the vector map and objects extracted from the input imagery.

A rather simple change detection technique is presented by Steinle et al. (1999). A 3D CAD model of the old state should be compared with laser scanner data representing the new state. A point-wise comparison of heights is carried out and in cases where the laser scanner points are above the 3D CAD situation a change registered. The quality of this analysis highly depends on the completeness and correctness of the 3D CAD model. In cases where objects, also vegetation, are not included, false changes will be detected.

All these examples show that LiDAR data is nowadays a common source for change detection and updating procedures [Savopol and Armenakis (2004), Vosselman et al. 2004], but also spaceborne remote sensing images have proven important for change detection analyses, for instance, for analysing changes after earthquakes [Sirinyildiz (2004)], or mapping and modelling urbanisation processes [Cabrel et al. (2004), Pantazis et al. (2004)].

3 Sensors and Data

This section introduces the data sets that are used in this research study. They come from a wide range of sensors, both spaceborne and airborne, of which some technical characteristics are discussed in the following.

3.a Spaceborne Digital Acquisition Systems

In satellite remote sensing the idea of using a digital sensor to collect ground data is quite old. By the late 1960s, the first unmanned satellite specifically dedicated to multispectral remote sensing entered the planning stages. It was in July 1972 when the Landsat1 satellite (originally called ERTS1) was successfully launched [McDowell (1997)].

Landsat1 was using a whiskbroom scanner, which was acquiring pixel by pixel on the ground in order to assemble a full digital image. The sensor with its 80m geometric resolution and its four spectral channels was dedicated to small scale multispectral analysis rather than to geometric object recognition. Also the geometric quality was too bad to be useful for precision mapping, advanced analysis for detailed urban, cadastral and infrastructure planning. DSMs, in today's sense, being capable to acquire artificial and natural objects that protrude the terrain's surface, were not possible to be produced with first-generation satellite sensors. For these applications areas only airborne Photogrammetry would provide acceptable solutions.

As the years passed, technology progressed, the satellites were equipped with more advanced telescopic optical lenses, and the signal-to-noise ratio of the sensors became much better, meaning that the sensor size could be reduced remarkably. By decreasing the noise-equivalent-reflectance change, minimal radiating intensities collected by the sensor still delivered a reasonable (noiseless) signal. These facts mainly promised a higher geometric resolution.

When in September 1999 the first commercial high-resolution satellite called IKONOS was launched, a new era in remote sensing analysis began [Fraser and Hanley (2001)]. IKONOS captures ground information with a GSD (ground sample distance) of one metre. About a year later (December 2000) the EROS-1A (2.5 metres) mission started, and in October 2001 the Quickbird satellite, which provides nowadays the highest resolution from space for commercial purposes (0.61 metres) was launched. It was in May 2002 when SPOT5, which captures data with a footprint of 2.5 metres, was successfully brought to orbit. Finally, in June 2003, Orbview3 was launched providing a GSD of one metre. These satellite sensors are going to replace small scale aerial photographs for certain applications.

Table 3-1 provides an overview over today's state of optical sensors [SPOT (2004)]:

Table 3-1: Today's state of optical satellites.

Satellite	Sensor / Mode	GSD (m)	Area (km^2)
Landsat 7	ETM+ (PAN) ETM+ (MS)	15 30/60	31,450 31,450
ASTER	VNIR SWIR	15 30	3,600 3,600
IRS 1C, 1D	PAN LISS 3	5.8 25/70	4,900 19,600
IRS P6	LISS 3 LISS 4 (PAN) LISS 4 (MS)	23 5.8 5.8	19,600 4,900 571
IKONOS	PAN MS	1 4	121 121
Quickbird	PAN MS	0.6 2.4	272 272
Orbview	PAN MS	1 4	64 64
SPOT 1,2,3	XS PAN	20 10	3,600 3,600
SPOT 4	XI PAN	20 10	3,600 3,600
SPOT 5	XI PAN	10 2.5	3,600 3,600

Unfortunately these high-resolution sensors are associated with a big problem. Due to the fact that they have new geometric properties, the common methods (based on central perspective geometry) as they are used in airborne photogrammetry, cannot be directly adopted, at least not without certain modifications. This topic is broadly discussed in [Vozikis et al. (2003)].

One of the two major shortcomings of this imagery is the narrow field of view (IKONOS: 0.92º), which makes it difficult or even impossible to use the well-known collinearity equation [Kraus (1997)] for data analysis. The narrow field of view causes intersection problems of the imaging rays [Vozikis et al. (2003)].

The second deficiency is the lack of calibration data: no information about the interior and exterior orientation parameters is given. For orienting the imagery the so-called Rational Polynomial Coefficients (RPCs) are used [Dial and Grodecki (2002)]. Through these parameters (over 80 for each image), a relation between object and image space is established. Control information may be reduced to one or two control points. These RPCs are usually delivered together with the high-resolution satellite imagery and are supported by most of the commercial image exploitation softwares.

In the following some more information is given regarding the sensors, from which, data was used for this research.

3.a.i QUICKBIRD (DigitalGlobe)

Figure 3-1: Quickbird

In October 2001 DigitalGlobe brought Quickbird (Figure 3-1) into orbit. With its panchromatic linear CCD-array it is able to quickly change its pointing direction in any direction. Changing the viewing angle from fore- to aft-direction while the satellite travels on its orbit allows the acquisition of so-called along-track stereo pairs. See also directory.eoportal.org/pres_QUICKBIRD2.html (last accessed on 20 May 2005).

3.a.ii IKONOS (SpaceImaging)

The IKONOS satellite (Figure 3-2) was launched in September 1999 as the first high-resolution satellite providing data for commercial purposes. The satellite travels at approximately 7 kilometres per second over the ground and collects data at a rate of over 2000 square kilometres per minute. IKONOS can also acquire stereo-images both, in across-track mode (observing the area of interest from a neighbouring orbit) and in along-track direction.
See also directory.eoportal.org/pres_IKONOS2.html (last accessed on 20 May 2005).

Figure 3-2: IKONOS

3.a.iii Orbview3 (OrbImage)

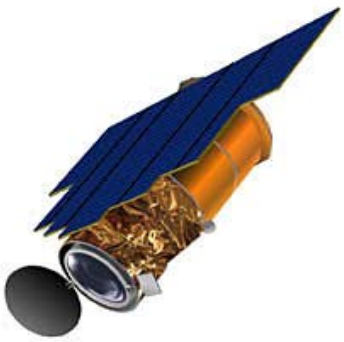

Figure 3-3: Orbview3

It was in June 2003 when the Orbview3 satellite started its operability (Figure 3-3). Equipped with a 0.45m aperture camera it takes 1m ground resolution images in the panchromatic spectrum and 4m in the multispectral. The geometric resolution characteristics are identical with the ones of the IKONOS satellite. A difference is the swath width which is 8km for Orbview3.

For more details see directory.eoportal.org/pres_OrbView3.html (last accessed on 20 May 2005).

3.a.iv SPOT5 (SPOT Image)

The SPOT5 satellite (Figure 3-4), which is equipped with three major instruments, was successfully launched in May 2002. Two of them are explained below.

The first one is the High Resolution Geometric (HRG) instrument, which is capable to generate data at a maximum resolution level of 2.5m covering a swath of 60km. The second device on board of the SPOT5 satellite is the High Resolution Stereoscopic (HRS) instrument, which has the ability to acquire stereo pair images nearly simultaneously. Only 1.5 minutes after the forward-looking telescope acquired images of the ground, with a viewing angle of 20° ahead of the nadir, the aft-looking telescope acquires the same strip. This is a great advantage for high-resolution DSM production. See also directory.eoportal.org/pres_SPOT5.html (last accessed on 20 May 2005).

Figure 3-4: SPOT5

3.b Airborne Digital Line Scanners

In the 1980s Otto Hofmann designed and started developing the first airborne line scanning sensor for acquiring stereo images, thus explicitly introducing this kind of sensor for purposes of object reconstruction rather than for pure interpretation tasks [Fritsch (1997)]. He designed a three-line scanning system with a forward-, backward- and nadir-looking linear sensor array thus delivering in-flight stereo views and he also thoroughly described an appropriate functional model for block adjustment. With

the spaceborne MOMS-2P sensor this idea was firstly put into practice [Kornus and Lehner (1999)].

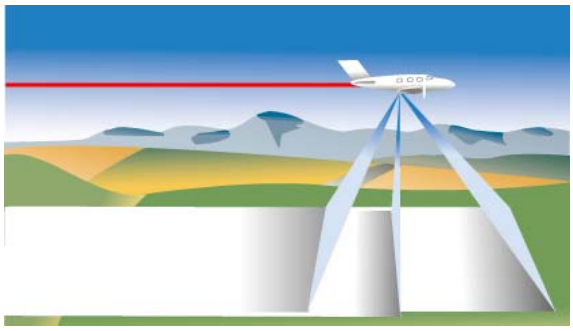

Figure 3-5: Principle of a three-line scanner.

The concept of the airborne three-line scanner was implemented in 1997 into the HRSC–AX047 (High Resolution Stereoscopic Camera) by DLR (German Aerospace Centre) [Neukum (2001)], and was then, in 2000, commercialized through the ADS40 (Airborne Digital Sensor) by LH-Systems [Fricker (2001)].

The big advantage when airborne scanners are equipped with multiple detector lines is the fact that they acquire information about the same area on the ground with multiple viewing angles nearly synchronously (Figure 3-5). This of course makes the deriving of DSMs simpler. Each point on the ground of the acquired area is depicted on three images at least, which of course increases the accuracy of the calculated DSM.

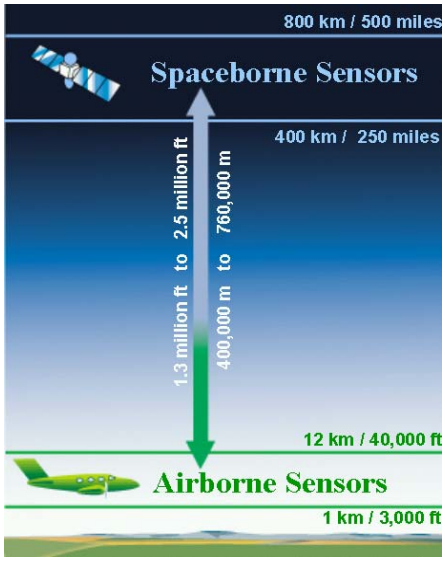

Figure 3-6: Spaceborne vs. Airborne.

The problem of the narrow field of view as it is found at high-resolution satellites does not exist here. Since the airplane flies very low (between 1000 and 10000

metres) in comparison to the satellites (400 kilometres or even more) (Figure 3-6), airborne line sensors use a rather wide field of view (ADS40: 64°), as we know it from the common analogue airborne frame cameras.

However there is a disadvantage when flying inside the atmosphere. Unfortunately the airplane does not have such a smooth flying path as satellites do. The flying path (Keplarian orbit) of a satellite can be described through an analytic mathematical function [Westin (1990)].

When looking at line scanning systems, each acquired line has its own yaw, pitch and roll. Since satellites follow a smooth orbit the exterior orientation parameters between neighbouring lines are highly correlated and therefore predictable. This is not the case when dealing with airborne line scanning systems mainly due to the unsteadiness of the airplane caused by wind, thermal effects etc. In order to calculate precisely the exterior orientation for each acquired line, airborne line sensors need to be equipped with a highly sensitive INS (Inertial Navigation System), which records the precise orientation angles of the camera during the acquisition. This unit works together with a differential GPS [Fricker (2001)].

Figure 3-7: DSM of Vienna (by courtesy of ISTAR).

Once the stereo pairs have been carefully oriented, DSM extraction methods can be applied, the surface model generated and if required rendered with spectral information. Figure 3-7 shows a DSM derived by the company ISTAR from data acquired with the HRSC-AX sensor.

A few details about the HRSC and ADS40 sensors are given in the next two paragraphs.

3.b.i HRSC-AX (DLR)

Since September 2000 the company ISTAR has been leasing the wide-angle version HRSC-AX047 (referred to as HRSC-AXW) from DLR to mainly acquire data over the US. The follow-up normal-angle model HRSC-AX150 (referred to as HRSC-AX) started being used by ISTAR from July 2001 mainly in Europe. Some data on the HRSC-AX150 (Figure 3-8):

- Focal length: 150 mm,
- Swath angle: 29.1°,
- 9 CCD lines: 12000 pixels each.

Figure 3-8: HRSC-AX

3.b.ii ADS40 (LH-Systems)


Figure 3-9: ADS40

The ADS40 (Figure 3-9) was introduced at the XIX ISPRS Congress in Amsterdam. Some general technical information:

- 3 panchomatic CCD lines, each 2 x 12000 pixels, staggered by 3.25 µm
- 4 multispectral CCD lines each 12000 pixels
- Pixel size: 6.5 µm x 6.5 µm.
- Swath angle: 64°
- Focal length: 62.77 mm

3.c Used Data

In the following the specifications of the test data will be presented which have been used for this work (Table 3-2). It should be mentioned here that only subsets have been used for the investigations.

Table 3-2: Examined data sets.

Sensor	Location	GSD (m)	Extents (km)	Stereo	DSM
ADS40	Valladolid, Spain	0.25	1 x 1	No	by ISTAR
ADS40	Nîmes, France	ca. 0.2	0.2 x 0.2	No	by DLR
HRSC-AX	Bern, Switzerland	ca. 0.3	0.2 x 0.3	No	by DLR
Quickbird	Phoenix, USA	0.6	16.5 x 12.6	Yes	with LPS
Quickbird	Denver, USA	0.6	16.9 x 16.5	Yes	with LPS
IKONOS	Athens, Greece	1	9.7 x 12.3	Yes	with LPS
Orbview 3	Orange, USA	1	0.6 x 0.7	No	
SPOT5	Attica, Greece	2.5	60 x 60	No	
SPOT5	Rome, Italy	2.5	60 x 60	No	
SPOT4	Attica, Greece	10	82.7 x 70.6	No	
Landsat 7	Neusiedlersee, Austria	30	30.8 x 30.8	No	

The Valladolid orthoimage (0.25m resolution) is a good example for high-resolution mapping by employing a digital airborne line scanning sensor. The corresponding DSM (1m resolution) gives a good impression on the exploitation possibilities of such imagery (Figure 3-10). Since the side overlap during this flight was extremely high (up to 80%) each point on the ground can be seen in at least nine images (or better to say image strips). In case of such a high redundancy, multi-ray matching techniques make it possible to extract DSMs with qualities that can easily compete with laser scanner products (especially in urban areas). The true orthophoto and the corresponding DSM are provided by courtesy of the company ISTAR (www.istar.fr, last accessed on 20 May 2005).

Figure 3-10: DSM with draped orthophoto (Valladolid, Spain).

The second ADS40 and the HRSC-AX data sets (Figure 3-11) were taken from Scholten and Gwinner (2003). Their goal was to automate the production of the corresponding DSMs and orthophotos as much as possible. This goal was achieved by a procedure realized at the DLR (www.dlr.de, last accessed on 20 May 2005), where the only information used for image orientation, matching and ortho-resampling was the

GPS and IMU data that was acquired during the flight. The introduced methodology is based on a network of computers that work on a highly autonomous level. The only human interaction is needed for quality control purposes.

Beware that the GSD mentioned in Table 3-2 for these two data sets differs from the one mentioned in Scholten and Gwinner (2003); this is due to the fact that pushbroom scanners usually have a different resolution in along-track and in across-track direction. The geometric resolution in across-track direction is a function of the flying height and the size of the CCD unit on the sensor plate, whereas the resolution in along-track direction additionally depends on flying speed of the aircraft as well as on the exposure time. It is tried to adjust the exposure time in such a way so that the footprint size is equal to the GSD and thus neither oversampling nor undersampling occurs. The GSD values stated in Table 3-2 are the mean of the along-track and across-track values.

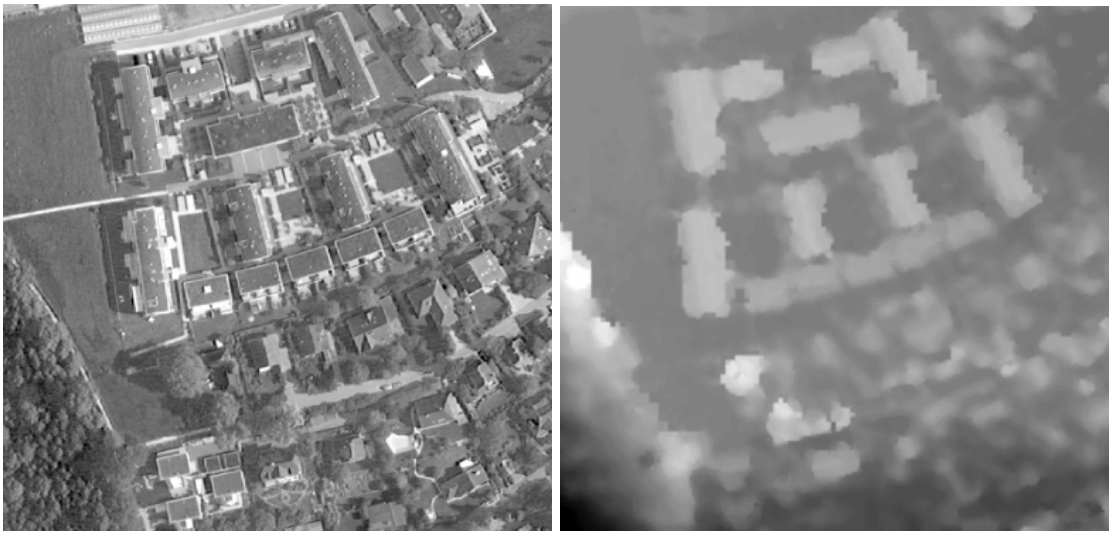

Figure 3-11: Subset of HRSC-AX nadir image and corresponding grey-level-coded DSM.

The two used stereo Quickbird imageries (Phoenix, USA and Denver, USA) were provided by courtesy of Eurimage (www.eurimage.it, last accessed on 3 March 2005). Both consist of a panchromatic stereo pair of images, and were delivered together with their RPCs. The DSMs were extracted by using the software package LPS version 8.7 by Leica Geosystems (www.leica-geosystems.com, last accessed on 20 May 2005). Since the covered areas are huge (ca. 150 km^2), buildings of various sizes may occur. E.g. on the one hand there are huge factory buildings and on the other hand small maisonettes can be found (Figure 3-12). This variability in size must be taken into account when performing the later described building extraction methodologies.

Figure 3-12: Quickbird subsets depicting differently sized buildings.

The IKONOS stereo pair depicts a region north-east of Athens, Greece and was made available by the courtesy of company IGD (www.igd.gr, last accessed on 20 May 2005). Here too, the RPCs were delivered together with the imagery and used for image orientation and DSM creation. Figure 3-13 shows clearly that if dealing with rather small objects, the quality of the derived DSM decreases tremendously, especially if compared to data of the matching results from aerial acquisitions (Figure 3-10). It is very difficult to recognize individual buildings in the DSM; but still, this coarse DSM information will prove to be useful for some building extraction and updating techniques.

Figure 3-13: Subset of the IKONOS image and the derived DSM.

The Orbview3 image depicts a small region of a suburb of Orange, USA and was taken from the image gallery of the official OrbImange website (www.obrimage.com, last accessed on 15 April 2005).

The four last data sets were not used directly for building extraction and updating, but provided useful information for testing the later discussed algorithms.

The SPOT4 and SPOT5 data sets of Attica, Greece are panchromatic images provided by the courtesy of company IGD, whereas the other SPOT5 (Rome, Italy) panchromatic image was taken from a demo CD-ROM of company SpotImage (www.spotimage.fr, last accessed on 20 May 2005).

The Landsat7 imagery is multispectral (7 bands) and has been provided as an educational data set from IPF (Institute of Photogrammetry and Remote Sensing) of Vienna University of Technology (www.ipf.tuwien.ac.at, last accessed on 20 May 2005).

4 Proposed Workflow

Automated building detection systems using imageries have evolved along many lines, but the trend has always been towards greater generality [Jaynes et al. (2003)]. Research has evolved from:

- Restricted nadir views [Nagao et al. (1979), Venkateswar and Chellapa (1990)]: this assumption greatly simplifies building extraction since rectangular building roofs appear as rectangles in the image.
- General oblique view points [Collins et al. (1998), McGlone and Schufelt (1994)]: rectangular rooftops might appear as parallelograms if weak-perspective or affine views are assumed.
- Single image analysis [Jaynes (2000)]: building height information through shadow analysis.
- Multi image techniques [Fua and Hanson (1991)]: get height information through aerotriangulation of two or more images.
- Purely 2D hypothesis extraction in image space [Tavakoli and Rosenfeld (1982)]: building extraction in hierarchical stages starting from a very generalized form of the building up to detailed description.
- Rigorous 3D geometric reconstruction in object space [Foerstner (1995), Haala and Hahn (1995)]: buildings are assumed to have similar heights.

The proposed workflow (Figure 4-1) for deriving DCMs promises higher flexibility and is affected by one restriction: the input data must either be stereo imagery or an additional elevation source must be available in order to reconstruct the 3rd dimension. The approach is conceived for approximate nadir view images, but some thoughts are described in the following how to deal with data captured with oblique viewing angles.

It can be subdivided into four major parts, namely pre-processing, nDSM creation, buildings localization and buildings extraction.

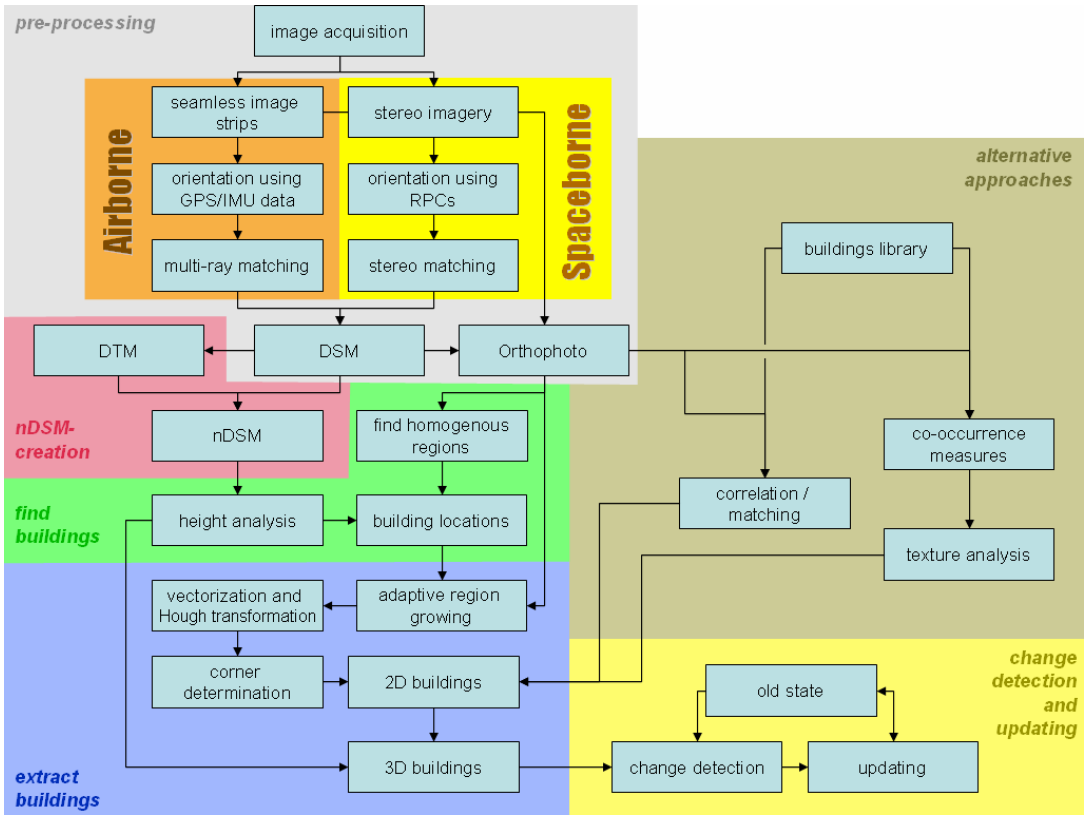

Figure 4-1: Proposed workflow.

4.a Pre-Processing

High-resolution satellites deliver overlapping image scenes in form of stereo pairs, whereas aerial multi-line scanners generate one image strip for each sensor line, where each strip is associated with a different viewing angle. The orientation of multi-line sensor data is automated to a very high degree.

4.a.i Orientation

Vozikis et al. (2003) have shown that images from high-resolution satellite sensors cannot be oriented by applying the common procedures used in aerial photogrammetry, because, firstly, most distributors do not provide any calibration information (interior orientation parameters) and, secondly, due to the extremely narrow field of view the bad intersection quality of the rays would cause the well-known collinearity equation to fail. Diverse orientation models have been proposed, but the fastest way is employing the so-called RPCs, which are usually provided together with the images. These coefficients describe the geometric relation between

object and image space so that the end-user can immediately start image exploitation. Unfortunately it turns out that the RPCs may be affected by a constant shift and hence they have to be refined by using two or three GCPs (Ground Control Points) in order to guarantee appropriate geometric quality [Hanley et al. (2002)].

For the airborne line scanner image strips, the orientation procedure is different. Together with the image data additional information from a GPS and IMU device are recorded at a very high temporal resolution (ca. 200 MHZ), providing information about the exact position (10-30cm) and orientation (less than 30") of the sensor at the time of acquisition. This so-called direct georeferencing is crucial, as due to the dynamic acquisition principle and due to the unsteadiness of the airplane movement each scan-line has its individual exterior orientation and a conventional orientation procedure would not (without significant effort) be able to deliver the required accuracy. The typical process of this direct georeferencing is described in Fricker (2001).
In the usual case of an image block, formed by several overlapping image strips, tie points measurement would follow in order to be able to perform an aero-triangulation. It is recommended to use some GCPs, at least at the block corners, but as shown by the good results from Scholten and Gwinner (2003) it is not always necessary since the collected orientation data from the GPS and IMU may be sufficient.

4.a.ii DSM Extraction

Once the images are accurately oriented (multi-ray) matching algorithms can be performed for deriving a DSM. Finally, the images and the DSM can be used to create an orthophoto of the scene, which can also be used as input for the building extraction algorithm.

Line scanner data cannot be processed in the same way as frame imagery (whether digitally acquired or scanned analogue film), but still it is a straight forward process; from acquisition to the DSM [Boerner et al. (1997)].

In Sandau and Eckhardt (1996) a processing chain is tested with the WAAC (Wide Angle Airborne Camera) CCD-line scanner. The major steps are:

- collection of image, positioning and attitude data
- geometric correction (using GPS and IMU data)
- get initial values for start-points for matching

- matching
- calculating the object coordinates of the matched points
- interpolation (DSM creation)

No user interaction is required during this process.

Also at the DLR a processing chain for handling airborne three line scanner imagery (HRSC-AX and ADS40) has been developed and automated to such a degree, that human interaction is only needed at the data acquisition step [Scholten and Gwinner (2003)]. Once the data is downloaded onto the computers the whole process from image correction and rectification, to orientation and triangulation, up to the DSM and orthophoto production can be done fully automatically.

It is obvious that the higher the image overlap, the more reliable are the extracted DSM results. Henriccson (1996) derives a DSM from fourfold overlapping colour images and Baillard et al. (1999) uses images with a sixfold overlap. These approaches take advantage of the high number of available images and thus reduce the troubles that arise when parts of the surface are not contained in some of the images due to occlusion or other visibility problems.

Since the DSM is a very important input source for the creation of DCMs, some theory on matching is discussed below.

4.a.ii.A General Remarks on Image Matching

Matching problems are so-called corresponding problems. Nearly all of the digital photogrammetric workstations support such matching algorithms nowadays. Commonly used approaches are area-based matching (the primitives are windows composed of grey values), feature-based matching (the primitives are features extracted in each image a priori) or structural techniques.

For readings on matching see Kraus (1996), Kreiling (1976), Ackermann(1984), Helava (1988), Foerstner (1982), Hannah (1988), Gruen (1985) and Wang (1998).

Generally speaking, algorithms now frequently use constraints on the primitives in order to find an optimal solution. The most common are:

- Epipolar constraint
- Uniqueness constraint

- Surface continuity constraint
- Order constraint
- Photometry constraint

The decision which constraint(s) will be chosen for an algorithm has a big effect not only on the quality of the result, but also on the time of calculation.

4.a.ii.B Area-Based Matching

Area-based matching is often called signal-based matching and is considered more traditional. For these methods we usually need good initial values for the unknown parameters. Each point to be matched is the centre of a small window of pixels in the reference image (template), and this window is statistically compared with equally sized windows of pixels in other (target) images. The measure of match is either a difference metric that is minimised, such as RMS difference, or a similarity measure that is maximised, such as mean- and variance-normalised cross-correlation. Area-based matching usually works with local windows. In the following some techniques are listed:

- Cross-correlation
- Least squares matching
- Adaptive least squares matching
- Multipatch matching
- Multipoint matching

4.a.ii.C Feature-Based Matching

This technique comprises two stages. Firstly the detection of interesting features and their attributes in all images and secondly, the determination of the corresponding features. The two stages are related to each other in the sense that the feature extraction and computation of their attributes must be such that the second stage of the correspondence is easy, not sensitive to errors and precise. Some methods for feature-based matching:

- Point extraction
- Isotropy and weight of features
- Edge extraction and matching
- Regions extraction

4.a.ii.D Structural Matching

Structural matching is also referred to as relational matching. It establishes a correspondence between the primitives of one structural description and the primitives of a second structural description. A structural description is defined by a set of features (primitives) and their interrelationships. Since structural matching techniques utilize not only image features but also topological and geometrical relations among the features to determine the correspondence, the image matching tasks can be fully automated without any initial estimates or very coarse ones.

4.a.ii.E Matching Problems

Problems encountered in automatic DSM generation [Gruen and Zhang (2003)]:

- poor or no texture
- distinct object discontinuities
- local object patch is no planar face
- repetitive objects
- occlusions
- moving objects, including shadows
- multi-layered and transparent objects
- radiometric artefacts, as specular reflections and others

After the matched point cloud has been created the surface can either be described as a triangular network via e.g. a Delaunay-Triangulation, or by interpolating over the point cloud in object space. Such interpolation methods can be polynomial interpolations, interpolation by surface summation, or a least squares interpolation (linear prediction) [Kraus (2000)].

4.b DSM Normalization

Firstly, a few thoughts on terminology are discussed. Since literature is not uniform, an explanation is given about what is meant when speaking of DSMs, DTMs, nDSM, DOMs (Digital Object Models), DEMs (Digital Elevation Models) and DHMs (Digital Height Models).
Figure 4-2 gives a good impression on the difference between DSM and DTM: the DTM represents the terrain and the DSM the highest point of the ground's surface.

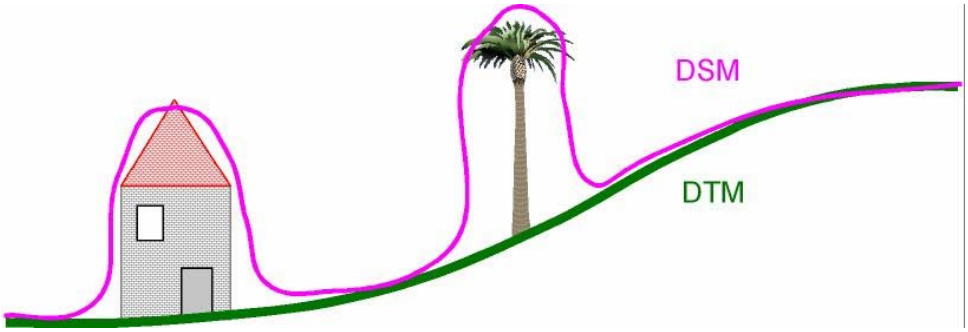

Figure 4-2: DSM vs. DTM.

Although the terms DEMs and DHMs are commonly used for the represention of the earth's surface (i.e. DTM), they may also stand for DSM.
The expression DOM is mostly found in close range applications for the model describing the object surface.
In some publications DCM is also referred to as Digital Canopy Model, which is equivalent to our DSM. It should not be mixed up with the DCM which is used throughout this work meaning Digital City Model.. For more details on the expressions relating to surfaces see Pfeifer (2003).

By DSM normalisation one understands the generation of a DSM of the off-terrain points. The DTM is then simply subtracted from the DSM, and the result is the nDSM (the normalised DSM), where terrain points have an elevation equal to zero and building points have their actual height above ground. Thus, all objects in the nDSM stand on elevation height zero. The crucial part of nDSM generation is, therefore, the generation of the DTM from a given DSM which process should be discussed in the following.

4.b.i nDSM Generation – State of the Art

As already mentioned, most of the approaches for normalizing the DSM aim at preliminary deriving a DTM.

A manual approach is described by Ehlers et al. (2003). They suggest a simple on-screen digitization of points in the DSM or orthophoto that represent the terrain's surface. These reference points are then given the corresponding height value from the DSM. Through these points an inverse distance weighting interpolation is calculated for computing the DTM.

The simplest automatic approach is to derive these terrain-reference-points by applying a local minimum operator on the DSM [Kilian et al. (1996)]. The kernel size of the minimum operator influences the quality of the DTM extraction. This procedure

might produce unwanted results in case the building sizes vary a lot and the kernel size cannot chosen adequately in advance. If the kernel size is chosen too small (i.e. smaller than buildings in the scene), points on top of the buildings will be included in the DTM. If the kernel size is chosen too large, important terrain details might be get lost.

Similarly, Baillard et al. (1998) suggest extracting above-ground blobs. It is checked if the blob height is superior than a critical value if compared to the heights of the neighbouring pixels. Thus the above-ground pixel is defined by its own height and by the height of the nature and of each neighbouring object. To be or not to be an above-ground area is a local and relative property based on differences in height between neighbouring segments of the DSM. Therefore symbolic representation of the scene as a set of homogeneous altimetric regions linked by neighbourhood relations is proposed. A graph is produced by segmenting the DSM into homogeneous altimetric regions. Each node is classified as ground or above-ground by means of Markovian modelling; neighbouring above-ground regions with similar heights are clustered into areas called above-ground blobs [Baillard et al. (1998)]. The remaining areas (ground-points) can then be used for deriving the terrain surface.

Another idea to derive the ground points of a DSM (for interpolating a DTM) is to apply a top-hat filter on the DSM [Weidner and Foerstner (1995)]. The top-hat operator is implemented by first applying the (morphological) opening operator on the original DSM image and then subtracting the result from the original image. Applying the top-hat operator provides a result that shows bright peaks in the image. These bright peaks correspond to above-ground points that will be excluded from the interpolation of the DTM.

Vosselman (2000) makes use of a morphological filter: it is an inverted funnel that searches for points beneath it. If there is no point beneath the inverted funnel found, the point at the centre of the cone is accepted as terrain point. Roggero (2001) continues this idea and adapts the shape of the inverted funnel to the slope of the terrain. The slope is determined by using a local linear regression criterion: each point is compared to the lowest point in the local operator neighbourhood. The distances and height differences are weighted and used as observations in the linear regression. Likewise, Sithole (2001) proposes to modify the inverted funnel and use as a local operator a cone. For better performance in steep terrain the cone angle is altered with the terrain slope.

Another approved method is to hierarchically filter the existing DSM-data by applying a skewed error distribution [Kraus (2000)]. To get the DTM, only points are used for the

interpolation that lie on the ground and not on buildings or vegetation. Usually this is done manually, but there exist some methods available to extract the DTM from the DSM by using all the matched points. The solution is provided by a certain filter, which uses gross error-information in order to give each point of the cloud an individual weight. These weights are distributed in such a way, that the points close to the ground influence the interpolation result more than the ones on buildings or vegetation. This procedure is called interpolation with skewed error distribution and algorithm is primarily applied on airborne laser scanner data [Briese (2002)] but can also be adopted on point clouds created via matching techniques [Bauerhansl et al. (2004)].

Elmqvist et al. (2001) employ a floating membrane that moves upwards from beneath the point cloud and clings to low points. Its shape is described by material characteristics like rigidity or elasticity etc. These parameters control the form of the membrane and the form of the terrain. Any point within a certain buffer of the membrane is accepted as terrain point.

An algorithm based on a two-step progressive densification of a TIN (triangular irregular network) is presented by Sohn and Dowman (2002). In this downward classification, four points closest to the rectangular bounds of the point cloud are chosen and triangulated. The lowest point of each triangle is added to the triangulation. This process is repeated for the triangles in the new triangulation until there are no triangles left with points beneath them. Points of the TIN are registered as terrain points. Since this way not all the terrain points are captured, an upward densification has to be carried out: a buffer is defined above every triangle and points within this buffer are tested using MDL (Minimum Distance Length) to find which gives the flattest tetrahedral. The points yielding the flattest tetrahedral are added to the triangulation.

Axelsson (1999) also performs an analysis on a TIN. A sparse TIN is derived from neighbourhood minima and then progressively densified to the laser point cloud. At each iteration step, a point is added to the TIN in case the angle that this point makes to the triangle is below a certain threshold and if it lies within a minimum distance from the nearest triangle node. After each iteration a new threshold is computed. This process is carried out until no more points are below the threshold.

Worth mentioning is also the algorithm discussed in Vosselman et al. (2004); its steps are described below:

- The point cloud is divided into sets of parallel thin slices in the xy-plane.

- Points of each slice are considered as a profile.
- A minimum spanning tree is computed for each profile.
- By removing the tree edges that exceed a certain slope or length threshold, the minimum spanning tree is segmented into line segments. All profiles are thus segmented. This procedure is repeated for other sets of profiles running in different orientations in the xy-planes.
- The resulting line segments of the different orientations are joint if they contain a common laser-point.
- The surfaces that are created have height discontinuities all around their contours.
- Advantage: this method can cope with multiple overlapping surfaces; both canopy points and ground points below are captured in different segments.
- Segments are classified based on the sign of height discontinuities at the ends of all line segments of a segment.
- Only segments with a low proportion of line segments that are above neighbouring line segments are classified as earth.

4.b.ii nDSM Generation in TABU

Most of the published algorithms aim at working with a highly dense point cloud acquired by airborne laser scanning and cannot be directly adopted to DSMs coming from matched airborne or spaceborne stereo imagery.

The procedure for generating normalized Digital Surface Models that is implemented in TABU can be divided into two steps. The main idea is the same as in most of the previously explained approaches: derive the DTM from the DSM and then subtract it from the DSM.

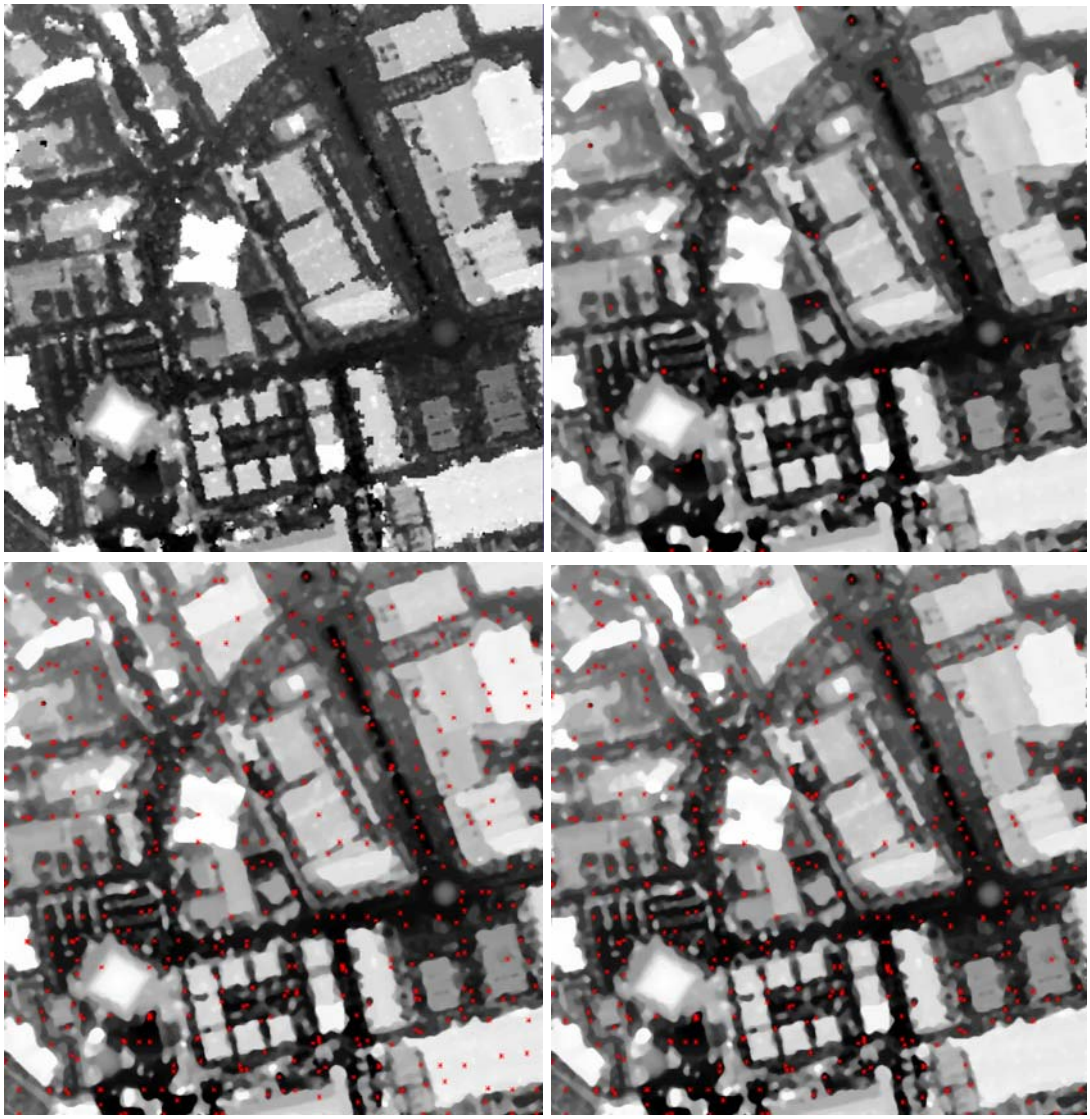

Figure 4-3: Top left: original DSM (ADS40-subset of Nimes); Top right: median-filtered DSM with about 75 computed terrain points; Lower left: calculation of approximately 300 terrain points without the big tiles; Lower right: same as lower left, but the big tiles window were employed too.

Our approach could be interpreted as an extension of the algorithm proposed by Kilian et al. (1996). The main difference is that not only one, but two local minimum operators are employed. The algorithm is very simple and straight forward:

- The DSM is tiled in xy-direction into big areas that are at least as big as the extents of the biggest building in the scene.
- For each of these tiles the minimum height is computed (BIG_{min}) and associated with the corresponding tile.
- Inside the big tiles, again a tiling is performed. The size of the small tiles depends on the wanted DTM detail.

- Inside each small tile the minimum height is searched ($SMALL_{min}$) together with its corresponding position ($SMALL_{xy}$)
- In case the difference between BIG_{min} and $SMALL_{min}$ does not exceed a certain threshold (=height of the lowest building in the whole dataset), $SMALL_{xy}$ is flagged as terrain point that will be used for DTM calculation.

Figure 4-3 shows clearly how arising gross errors, namely terrain points on top of buildings, for a small kernel size can be eliminated by using the previously explained two-step algorithm.

After the terrain points have been found an interpolation through these points is still pending in order to get the DTM. In TABU multiple interpolation methods are implemented for this task (Figure 4-4).

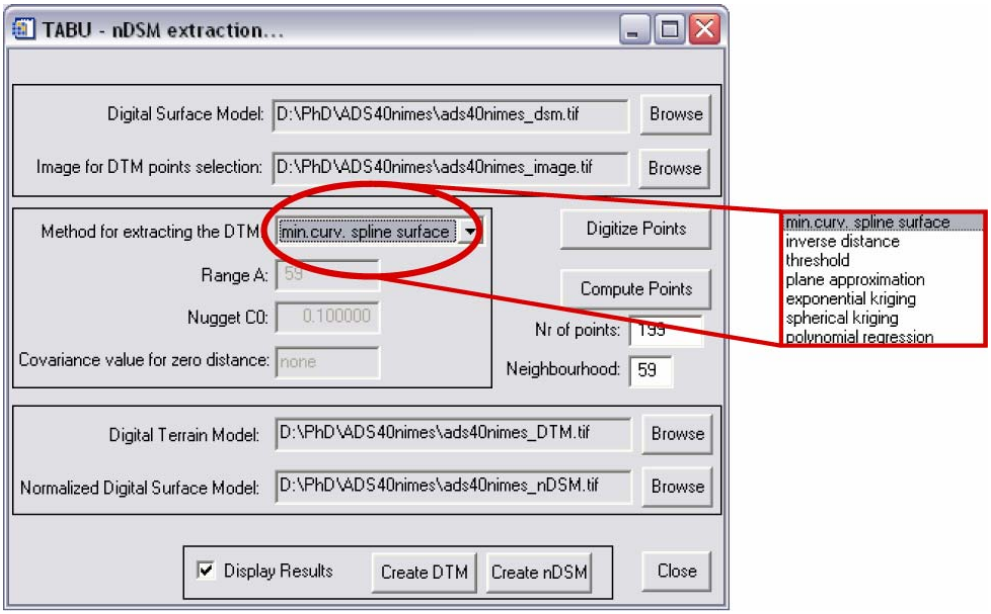

Figure 4-4: Screen dump of the parameter window for the nDSM creation showing possible interpolation strategies.

For most of the data sets the upmost interpolation strategy is chosen (minimum curvature spline surface), since it is faster than the kriging, regression and inverse distance approaches.

4.b.iii nDSM Generation – Example

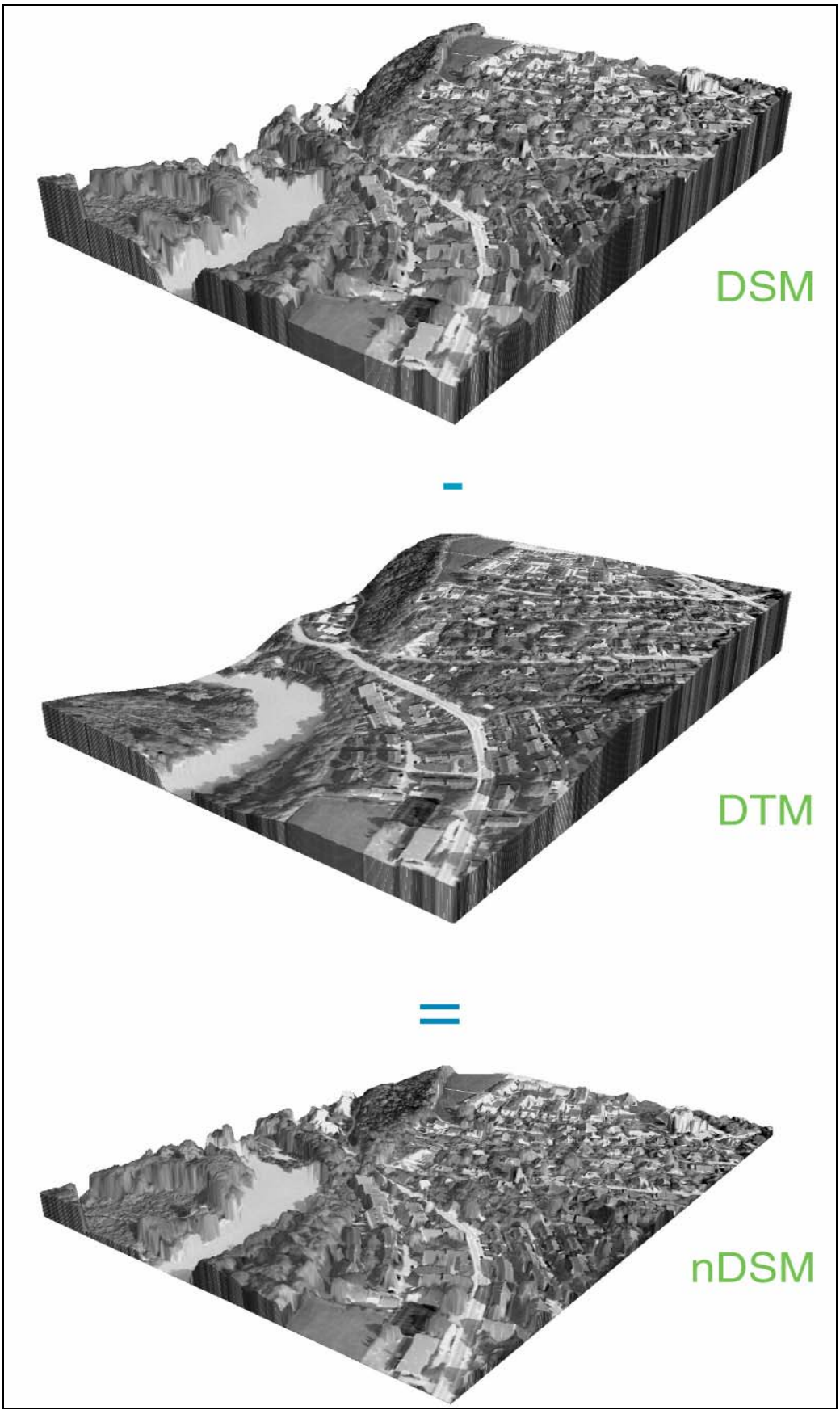

Figure 4-5: DSM, DTM and nDSM of the HRSC data set of Bern with draped orthophoto.

4.c Building Detection

The next and maybe most critical step is to find the regions of interest, i.e. the regions of potential candidates for buildings. The overall idea is to isolate buildings from the rest of the scene and place so-called seed points inside the buildings areas. These seed points are an important input for the next processing step, the Building Extraction.

Adams and Bischof (1994) describe a "seeded region growing" method, which is controlled by choosing a small number of pixels (seeds). The goal is to automatically find a tessellation of the image into regions, with the property that each connected component of a region meets exactly one of the predefined seeds. Neither training nor tuning parameters are needed, just the number of seed points. Hence this method is not applicable on imagery where "building seed points" have to be found, since we do not know in advance how many buildings are on the scene.

If multispectral information (especially NIR) were available, it would be easier to separate buildings from vegetation (or other objects) by applying a multispectral classification. But high resolution satellites provide high-resolution data only for the panchromatic bands. Hence a strategy has to be found to carry out this step, without using colour information.

The easiest approach would be to create a "building mask" using a certain elevation threshold on the nDSM [Rottensteiner (2003)]. This mask would point out areas where objects higher than the threshold are located, and everything below that height is considered as "non-building". "Non-building" can be low vegetation, cars or other objects positioned on the terrain surface, but not being high enough to be treated as building candidates. Unfortunately, it is not seldom that vegetation, especially trees or even forests, have the same height as buildings Figure 4-6 shows such a typical case. One can easily recognize that the dense trees to the left are at least as high as the buildings. In such a case simple thresholding would not make it possible to distinguish between objects of our interest (i.e. buildings) and vegetation.

When dealing with highly dense point clouds this problem can be solved quickly, since we expect to have rather heterogeneous areas over trees or bushes, in contrary to building areas where homogeneous surfaces are expected. Furthermore, the difference between first and last pulse of laser scanner data (if available) gives also a clear indication whether trees or buildings are on site.

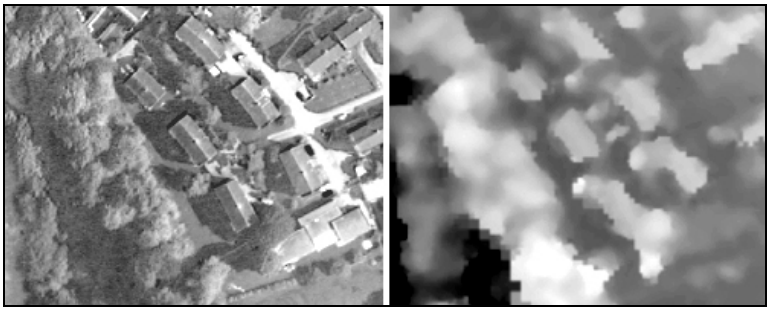

Figure 4-6: Image and nDSM subset of the HRSC-Bern data set.

Our data sets show that when dealing with DSMs derived from airborne or spaceborne data (e.g. Figure 4-6), the approach by analyzing roughness parameters of the DSM cannot be used for distinguishing between buildings and vegetation. This is why we utilize the radiometric information from the oriented images or orthophotos.

A thorough discussion on state of the art approaches for separating buildings from the rest of a scene can be found in section *2.a.i Segmentation of Image Data*.

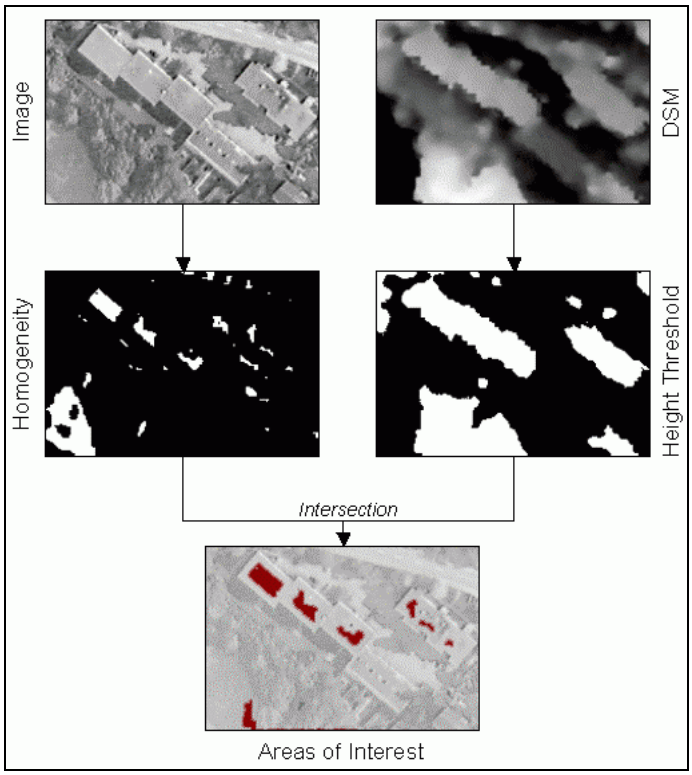

Figure 4-7: Finding building areas.

The algorithm proposed in this study for finding potential buildings regions is the following (see Figure 4-7):

- Set a certain elevation threshold of minimal building-height according to the current scene, then
- mark the potential candidate-regions of the nDSM in the image and
- apply a check for homogeneity in the image to refine the candidate selection.

This makes sense, since most of the times building-roofs are very homogeneous in the images e.g. one colour, or one regular pattern of tiles, in comparison to treetops or other vegetation. As mentioned before, to use the first derivative of the nDSM as a homogeneity parameter (a very popular method for high-resolution laser scanner data) is not recommended, because the derived nDSMs from our test data are much too coarse.

This specified homogeneity parameter can be calculated by looking at the statistical information of neighbouring pixels inside a certain kernel area, e.g. Variance, Minimum-Maximum-Difference, Skewness, Mean-Absolute-Deviation etc. (see Equations 5-5 to 5-16).

We are basically facing two problems of parameter setting: Firstly, defining an adequate size of the neighbourhood for which the homogeneity value is calculated, and secondly, determining a proper threshold that discriminates homogeneous from non-homogeneous areas. The example in Figure 4-7 shows two errors: one building was not found and one found area is not a building. Unfortunately, the success rate of this processing step not only depends on the chosen thresholds and kernel sizes, but also on the DSM quality. For the building that was not found, most probably the kernel size was chosen too big. But the other error, the wrongly determined building, is due to the inaccuracy of the nDSM in that region.

4.c.i Building Detection - Examples

The following two examples illustrate the proposed strategy for determining seed points. The actual seed points are placed inside found building areas of a certain size.

Figure 4-8 shows two different results for different thresholds. The first height threshold is chosen rather high, hence small buildings are neglected (e.g. the seven white buildings to the right of the image); altogether twelve seeds are determined (outcome 1). Then the threshold is decreased, so that the seven white buildings are included; altogether 23 seeds are derived (outcome 2). As a negative side effect

(when decreasing the threshold) also non-building areas are falsely characterized and seeds are set inside them. It turns out to be quite difficult for the algorithm to generate only correct results in areas where both, tall and low houses are situated.

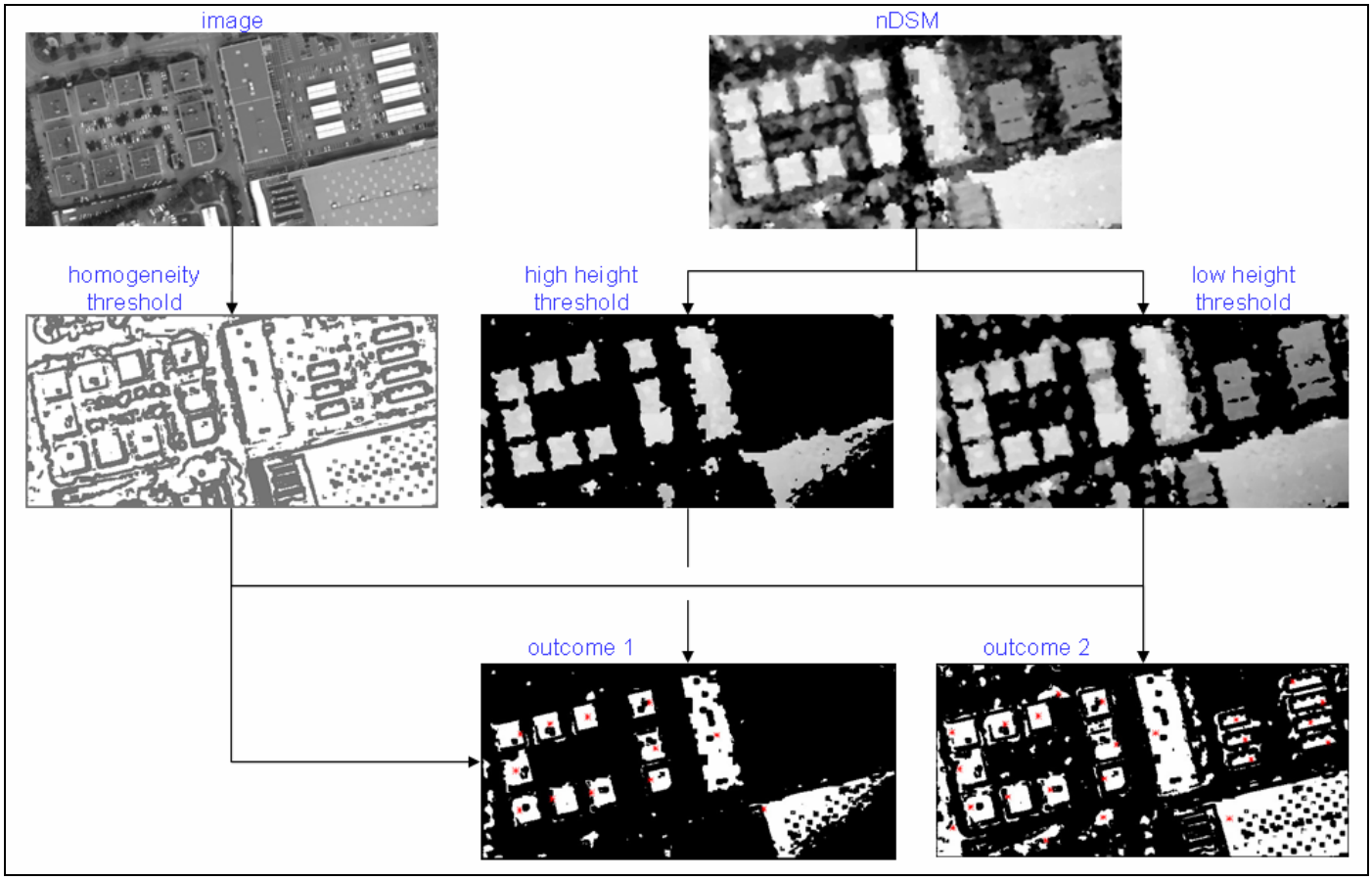

Figure 4-8: Determination of seed points with varying height thresholds (ADS40-subset of Nimes).

In the example of the HRSC-subset of Bern (Figure 4-9) it is shown that the homogeneity image is used to separate areas of the nDSM-threshold image that are falsely connected. For example, the lower horizontal building group; these five buildings are covered by one big blob in the nDSM (and the height thresholded nDSM). In contrary, in the homogeneity image these buildings are clearly separated due to the fact that building edges are heterogeneous in the imagery. Hence, in the final intersection image the buildings are clearly separated.

Concluding it can be said that the image information is not only important to distinguish between vegetation and buildings, but also to refine too coarse areas in the threshold-nDSM.

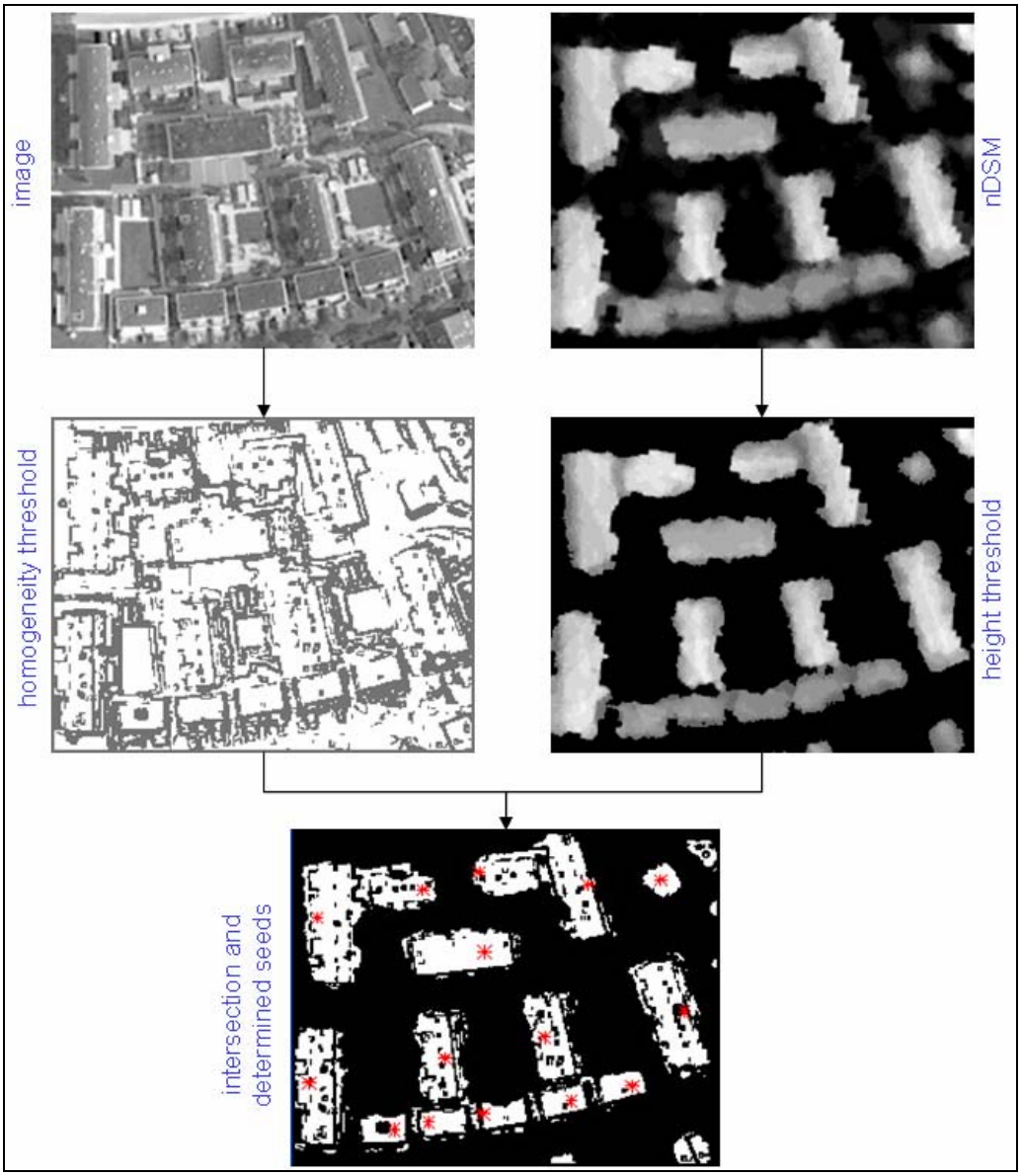

Figure 4-9: Seed point determination (HRSC-subset of Bern).

4.d Building Extraction

Once the regions of building candidates have been found and the seed points have been placed, the actual building extraction process can begin. In this step the goal is to ascertain the geometric properties of each building, in which a seed point was set. The building corners are stored in 3D in a data base, which comprises the major component of the DCM.

The coarseness of the used nDSMs does not allow the determination of the roof shape. Even more, it is not possible to get the building corners from the nDSMs (see Figure 4-7). Therefore additional information is taken from the oriented images or orthophotos. Since they provide higher geometric quality and textural information

they will be used to obtain the locations of corners and hence the shape of the buildings.

4.d.i Adaptive Region Growing

This section deals with the finding of all pixels belonging to a certain seed point. It could be interpreted as a segmentation step that separates the building under investigation from the rest of the image by finding all pixels that lie inside that building.

For the segmentation in intensity images four approaches are common [Haralick and Shapiro (1985)]:

- threshold techniques: pixels whose values lie within a certain threshold belong to one class (spatial information neglected, problems with blurred boundaries and noise)
- boundary based methods: rapid change of pixel values at the boundary between two regions (detection by applying a gradient operator)
- region-based methods: compare one pixel to its neighbours and check a certain homogeneity criteria (noise is a big problem)
- hybrid methods: combination of boundary and region criteria (includes morphological watershed segmentation and variable-order surface fitting)

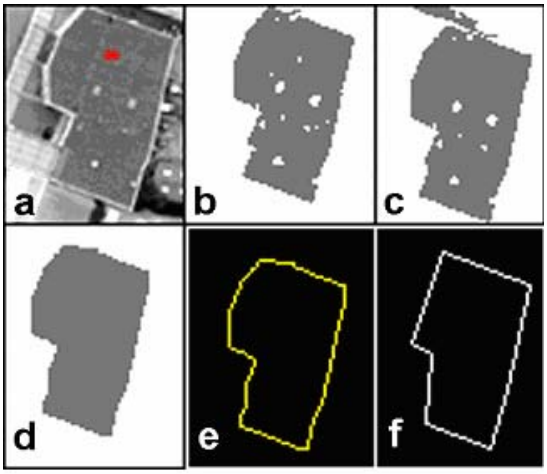

Figure 4-10: Example of building extraction.

We suggest an "adaptive" region-based method. The growing process of a certain region depends on assigned thresholds (see also Figure 4-28). The whole procedure is described in the following steps:

1. Inside the potential building start from the seed point with the region growing process and determine the parameters of the geometric object modelling the data of the seed region (Figure 4-10a, seed point=red asterisk).
2. Increase the threshold of region growing.
3. Apply region growing again and recalculate the geometric object parameters (Figure 4-10b).
4. If the calculated geometric object parameters differ largely from the geometric object parameters of the previous iteration step, or if they do not fulfil certain conditions, stop the iteration process, else go to step 1.

The geometric object parameters of the newly grown region can be:

- Area of the region,
- Perimeter of the region,
- Compactness $C=(4A\pi/P^2)$, where P is the perimeter and A the area [Kim and Anderson (1984)]. The range of C is between 0 and 1, the latter indicating the most compact region possible,
- The ratio of bounding box area and object area,
- Semi-major and semi-minor axis of the bounding ellipse [Post (1984)], or the ratio of them.

If these parameters vary a lot between two iteration steps, one can assume that the newly grown region went over the building edge as shown in Figure 4-10c. The conditions mentioned in step 4 can be used as additional information to stop the iteration process (For example the maximum extents of a building in our study area, or maximal number of corners).

Additionally, after every region growing process a sequence of blow and shrink functions (open-close) is applied onto the found regions (Figure 4-10d). This is necessary to fill holes that might exist within the building regions, or to smooth frayed out areas. The kernel size of the blow and shrink procedure must be chosen manually and it is recommended to choose a sequence of blowing and shrinking steps with individual kernel sizes (e.g. blow 3x3 – shrink 5x5).

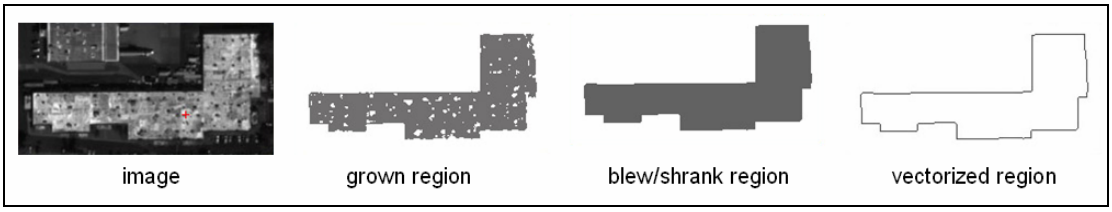

Figure 4-11: Example of the region growing, blow/shrink and vectorization process.

4.d.ii The Hough Transformation

The found region of interest is now vectorized (Figure 4-10e, Figure 4-11). The result is a very long coordinate list, since every single point along the building boundary is included. In the following we refer to this data as vectorized regions/buildings.
Many of the boundary points do not really represent the building geometry, but were created due to inaccuracies of the region-growing process. The task is to minimize this point list without impairing the geometric characteristics of the building (Figure 4-10f).

One solution would be to calculate regression lines through points of the vectorized data that possibly belong to the same edge of the building.

Another approach would be to make use of the Hough transformation. The general idea is to transform the information in the image (feature space) into a parameter space and apply there an analysis. It is a technique for isolating features that share common characteristics.
The classical Hough transformation is used to detect lines, circles, ellipses etc., whereas the generalized form can be used to detect features that cannot easily be described in an analytical way.

A line can be described through Equation 4-1 in image (feature) space. Now consider a space where not slope m and intercept value b are the parameters and x,y the coordinates (Equation 4-1), but the other way around: x,y are the parameters and m,b are the coordinates (Equation 4-2). This space is called parameter space or Hough domain.

$$y = mx + b \tag{4-1}$$

$$b = -xm + y \tag{4-2}$$

Figure 4-12 illustrates the relationship between the two spaces. A point in image space corresponds to a line in parameter space and the other way around.

Here, in each feature space four points exist that lie on an edge. These points correspond to lines in parameter space. The intersection points of these four lines in parameter space can be transformed back into feature space where they are defined as lines on which the original feature space points lie.

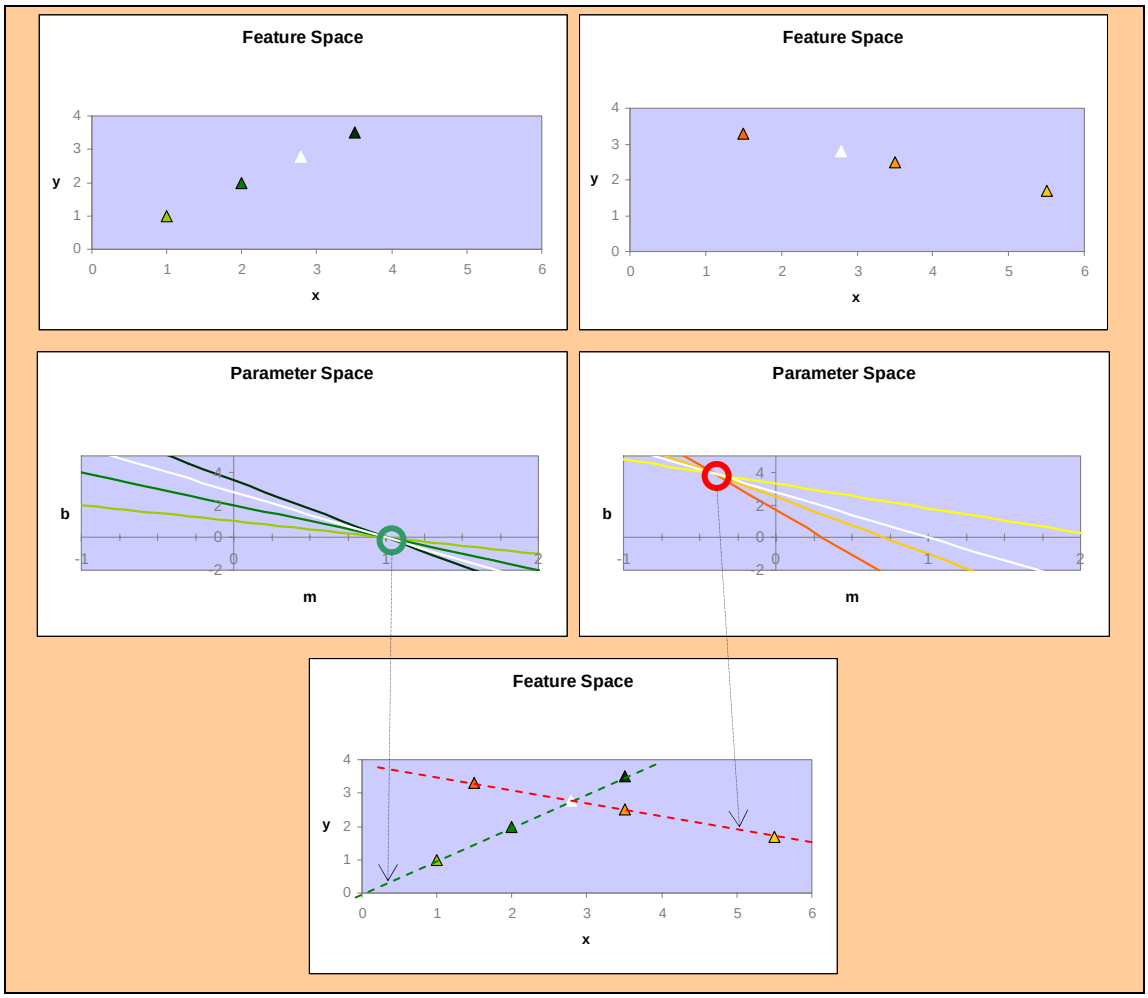

Figure 4-12: Illustration of relationship between feature (image) space and parameter space.

If the original points in feature space do not lie exactly on one straight edge, the corresponding lines in parameter space will not intersect in exactly one point.

It is obvious that these mathematical representations (Equations 4-1, 4-2) are afflicted with problems in case lines get very steep or even worse, if they get vertical. A vertical line in image space (feature space) means that the steepness m becomes infinite and hence cannot be transformed to and represented in parameter space. To avoid this drawback a different analytical line description, namely the Hough transform, is chosen (Equation 4-3).

$$\rho = x\cos(\theta) - y\sin(\theta) \tag{4-3}$$

where ρ is the perpendicular distance of a line from the origin
and θ the angle (in the range 0 to π) as illustrated in Figure 4-13.

To apply this function on the whole image, Equation 4-3 can be extended as shown in Equation 4-4 [IDL (2004)].

$$H(\theta,\rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x,y)\delta(\rho - x\cos(\theta) - y\sin(\theta))dxdy \tag{4-4}$$

where δ is the Dirac delta-function. Each point (x,y) in the original image $F(x,y)$ is transformed into a sinusoid $\rho = x\cos(\theta) - y\sin(\theta)$ (see Figure 4-13).

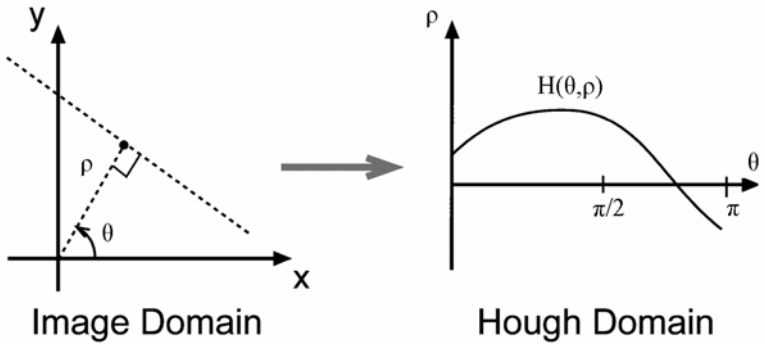

Figure 4-13: Hough transformation.

Points that lie on the same line in the image will produce sinusoids that all intersect at a single point in the Hough domain. For the inverse transform, or back-projection, each intersection point in the Hough domain is transformed into a straight line in the image.

Usually, the Hough function is used with binary images, in which case $H(\theta, \rho)$ gives the total number of sinusoids that intersect at point (θ, ρ), and hence, the total number of points making up the line in the original image. By choosing a threshold T for $H(\theta, \rho)$, and using the inverse Hough function, the original image is filtered to keep only lines that contain at least T points.

The computational attractiveness of the Hough transform arises from the subdividing of the parameter space into so-called accumulator cells. This would correspond to the discrete formula for the Hough transform (Equations 4-5, 4-6, 4-7).

$$H(\theta,\rho)=\sum_{i=0}^{I-1}\sum_{j=0}^{J-1}F_{ij}\delta(\rho,[\rho']) \tag{4-5}$$

$$\rho'=(i\cdot\Delta x+x_{\min})\cos\theta+(j\cdot\Delta y+y_{\min})\sin\theta \tag{4-6}$$

where F is the original image with array indices $i=0.....I-1$ and $j=0.....J-1$,
the brackets [] indicate rounding to the nearest integer, and
Δx and Δy are the spacings of the accumulator cells in the x and y directions.

The delta-function is defined as:

$$\delta(\rho,[\rho'])=\begin{cases}1 & \rho=[\rho'] \\ 0 & otherwise\end{cases} \tag{4-7}$$

Initially, all the accumulator cells are set to zero, and they are added the value 1 each time part of a transformed sinusoid lies inside them. Thus after all sinusoids have been projected the intersection points can be found just by searching for maximum values of the accumulator cells.
Another advantage is the fact that the computational accuracy can be adjusted easily just by varying the cell size Δx and Δy. By reducing the spacings the accuracy becomes better, but also computation time is increased.

The discrete formulas for the back-transformation are straightforward and for completeness given below (Equation 4-8) [IDL (2004)].

$$B_{ij}=\begin{cases}\sum_{\theta}\sum_{\rho}H(\theta,\rho)\cdot\delta(i,[m\cdot j+b]) & |\sin\theta|>\frac{\sqrt{2}}{2} \\ \sum_{\theta}\sum_{\rho}H(\theta,\rho)\cdot\delta(j,[m'\cdot i+b']) & |\sin\theta|\le\frac{\sqrt{2}}{2}\end{cases} \tag{4-8}$$

where m is the slope (Equations 4-9, 4-11) and
b the offset (Equations 4-10, 4-12) of the computed line in feature space.

$$m=-\frac{\Delta x}{\Delta y}\cdot\frac{\cos\theta}{\sin\theta} \tag{4-9}$$

$$b = \frac{\rho - x_{\min} \cdot \cos\theta - y_{\min} \cdot \sin\theta}{\Delta y \cdot \sin\theta} \tag{4-10}$$

$$m' = -\frac{\Delta y}{\Delta x} \cdot \frac{\sin\theta}{\cos\theta} \tag{4-11}$$

$$b' = \frac{\rho - x_{\min} \cdot \cos\theta - y_{\min} \cdot \sin\theta}{\Delta x \cdot \cos\theta} \tag{4-12}$$

Figure 4-14 depicts again the previous example, but this time the transformations between feature and parameter space were carried out by using Equation 4-4.

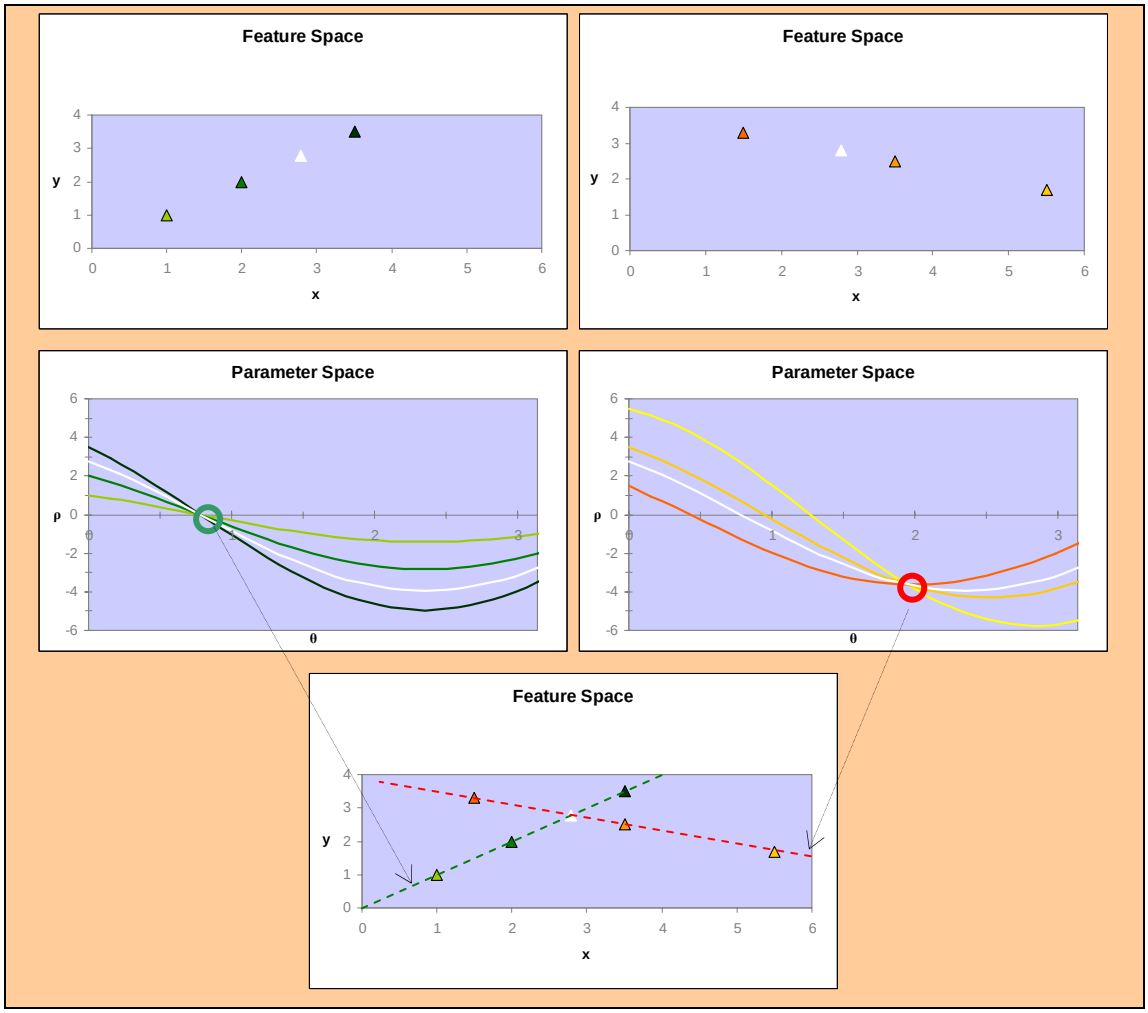

Figure 4-14: Illustration of relationship between feature (image) space and parameter space.

For further reading on the Hough Transformation see Jain (1989), Weeks (1996) Gonzalez and Woods (2002), Kraus (1990).

The appropriate resolution of the Hough domain can be derived from the input data:

- Vertical direction: The resolution of ρ is defined twice as high as the geometric resolution in feature space. This way the sampling theorem is fulfilled and thus edges in feature space separated by a distance of one pixel can be modelled separately.
- Horizontal direction: The maximum extent m_e of the investigated region is calculated. The resolution of θ is defined twice as high as the maximal allowable rotation of an edge of length m_e, so that the displacement of its endpoints is not bigger than one pixel.

4.d.iii Edge Detection (Hough Back-Transformation)

The problem we face is that points representing a building edge do not lie exactly on one single line, but close to it. This means that in Hough domain we have to look for a distribution of intersection points rather than an intersection of sinusoids in one single point. Investigations show that this distribution is not normal (Figure 4-15). Its shape depends on the scattering of the corresponding points around the edge in feature space. Goal is to determine the maximum of this distribution, which will be referred to in the following as local maximum.

The arising question is: To what a degree does this local maximum represent the centre of the distribution and hence the best fitting line in feature space?

We define the centre of the distribution as the weighted mean of all intersection points in Hough domain that are bigger than a predefined threshold; in our case 15 proved to be an adequate value. This means that only edges (in image space) are considered on which more than 15 points lie. As weight for each intersection the corresponding value from the discretized Hough domain is taken.

Six data sets of noisy lines are randomly generated. They consist of a sequence of points scattered along both sides of a straight line, thus simulating building edges after adaptive region growing. Also the corresponding sinusoids of these points are calculated. The distribution of the intersection points is visualized in Figure 4-15.
For each data set the locations of local maximum and weighted mean are calculated. Obviously, the worse the line is defined in image space the bigger is the deviation between the locations of local maximum and weighted mean.

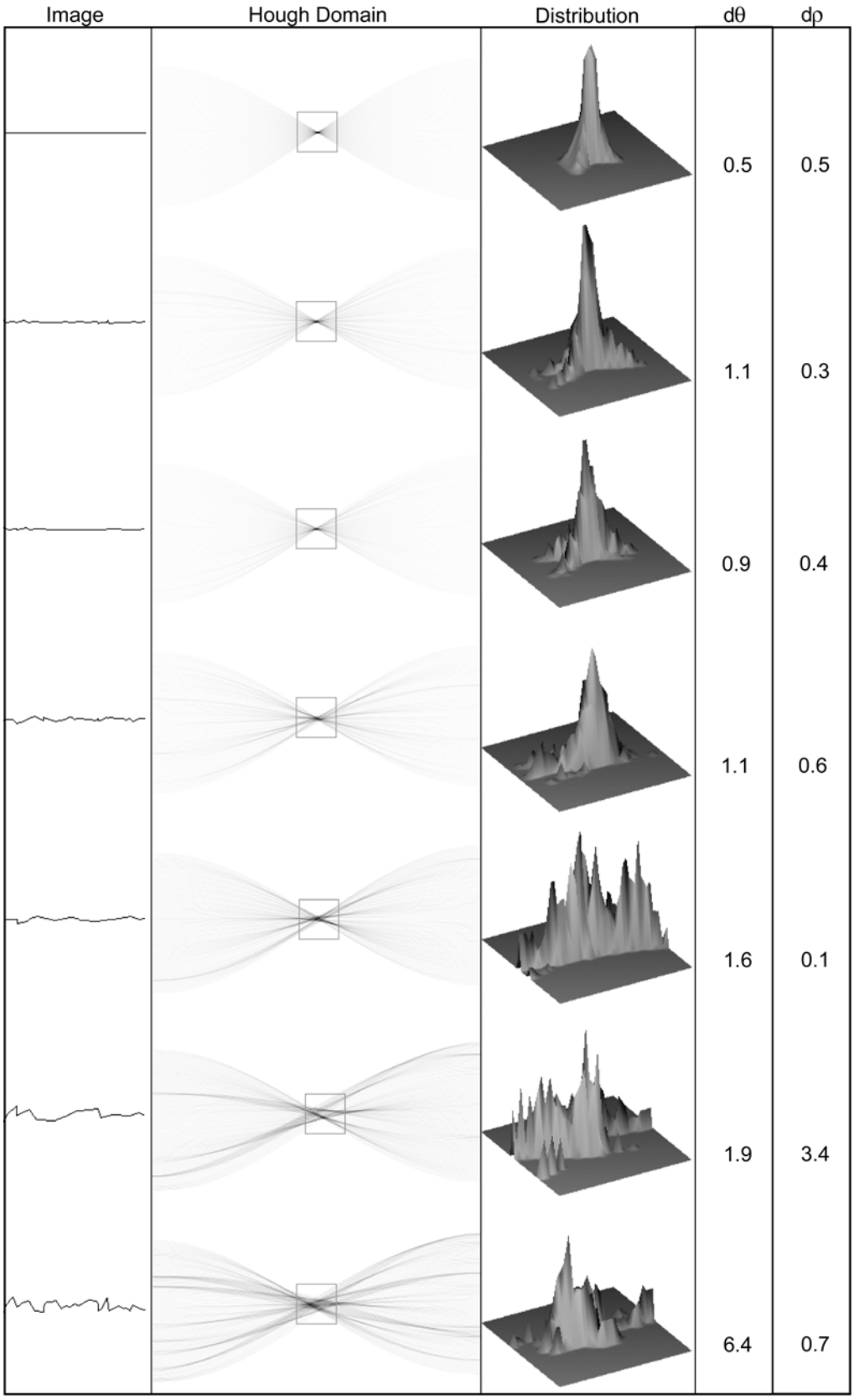

Figure 4-15: Distribution of sinusoid intersections ($d\theta$ [pixels] and $d\rho$ [pixels] are the absolute differences between the local maxima and the weighted means of the distributions in pixels in Hough domain).

Once the local maximum has been determined we can perform the back-transformation to obtain the according, representative line in image space. Unfortunately, the finding of local maxima (by employing a simple local maximum operator) in Hough space is quite tricky since the sinusoids are periodic function, which can lead to double-identifications close to the beginning ($\theta=0$) and end ($\theta=2\pi$) of the period.

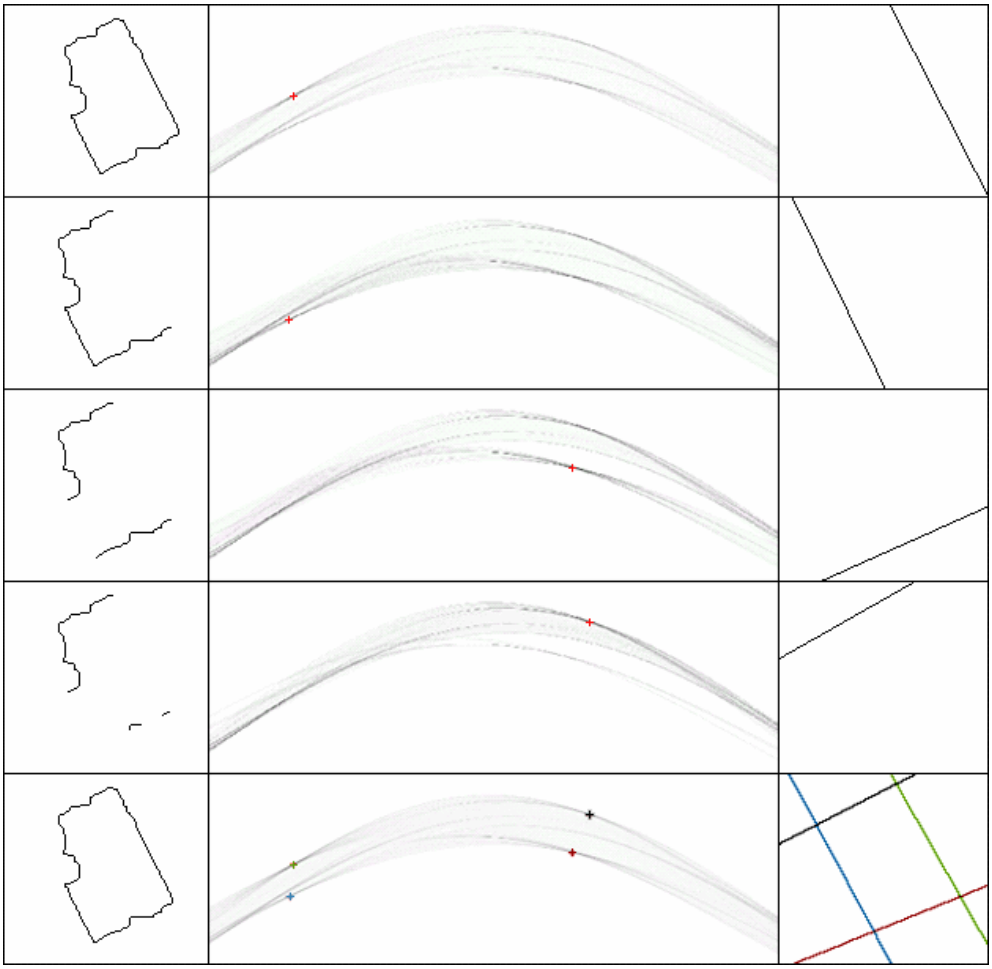

Figure 4-16: Iterations (=rows) of Hough transform (left column: vectorized building; centre column: Hough domain; right column: lines of back-transformed maxima)

The implemented solution to this problem carries out an iterative step-by-step Hough (back-) transformation (see Figure 4-16).

- First step: apply the Hough transformation to all points lying on the boundary of the found building.

- Start of iteration loop: Determine the global maximum in the Hough domain. This maximum location (=red cross) delivers exactly one θ and one ρ value.
- These two values represent one back-transformed line in image space. (Figure 4-16, first row).
- Eliminate all points in the vectorized image that have a smaller perpendicular distance than the tolerance τ from the back-transformed line. (Figure 4-16, second, third and forth row)
- Apply the Hough transformation again to the remaining image points and start next iteration.

This iterative process is terminated either if a requested number of lines has been found (= number of iteration steps), or when no more points lie on/near to a line. This would mean that the value at the found global maximum in Hough domain became smaller than the given threshold T.
In the example shown in Figure 4-16, the process was terminated after four iteration steps, which means that the building is described by four lines (Figure 4-16, last row).

Summing it up, the user enters to the system either how many edges have to be found for each building (e.g. if four is chosen, the buildings will have a quadrangle form), or the minimum number of points lying on an edge. In the latter case the buildings will be represented in more detail, but one has to bear in mind that the additional detail might be wrong coming from vectorization or region growing errors. Furthermore, the user also has to choose how big the tolerance value τ should be. This value depends on the quality of the vector model and hence the quality and resolution of the image.

4.d.iv Corner Determination

The intersection of the obtained lines can lead to wrong points that do not lie on the building boundary. To avoid this problem it has to be searched only for intersections that lie close to the vector data.

- Each back transformed edge is marked with a unique attribute (Figure 4-17c).
- All lines (and intersections) are dilated by the tolerance value τ (Figure 4-17d).

- Starting at a point on the vectorized building (Figure 4-17b) and moving along the edges in counter-clockwise direction, each pixel position is checked for the attribute of the closest dilated line.
- As soon as we have a change in line-attributes we search for the closest intersection of back-transformed edges and store the position as corner point (Figure 4-17e)!
- The procedure stops as soon as the start-position is reached.

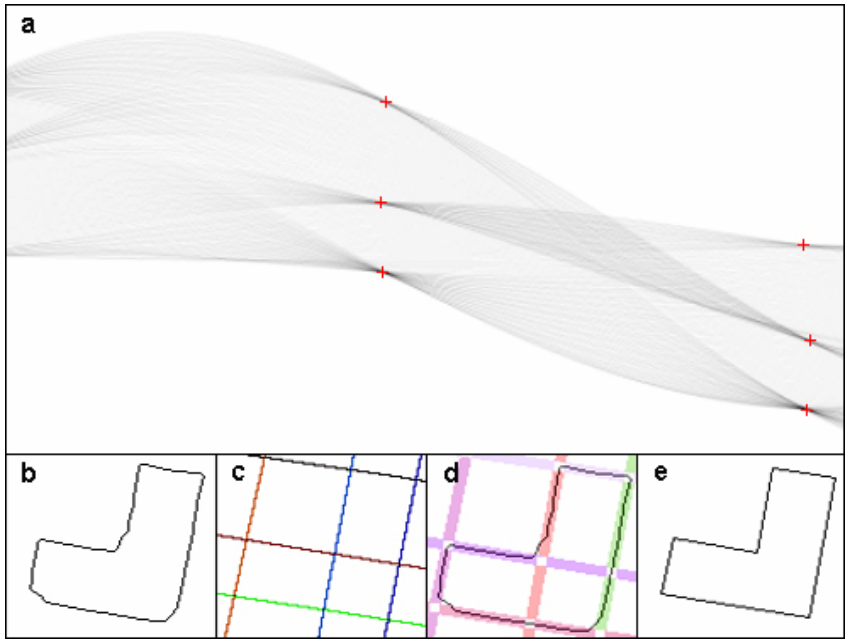

Figure 4-17: Finding the building corners (a: Hough domain with 6 determined maxima; b: vector representation; c: back-transformed lines; d: dilated back-transformed lines with vector data superimposed; e: final building represented by 6 points)

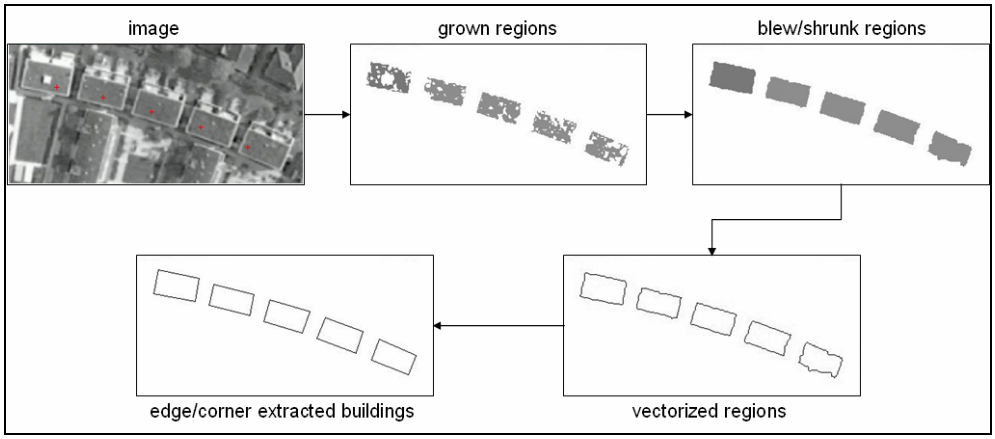

Figure 4-18: Example: building extraction (HRSC-subset of Bern).

Each extracted building will be assigned one height value (horizontal roof). This height value can be e.g. the median, mean or maximum of all points in the DSM lying within the investigated building boundaries. The building is completed by assuming vertical walls that are intersected with the DTM.

An analysis is carried out to investigate the reliability and quality of the chosen DSM heights for each building. Depending on the DSM resolution and accuracy, but also on the true roof shape, the elevation values that are used may vary. If we had a perfect DSM, the standard deviation of these values over a horizontal roof would yield zero. If a building in reality has a tilted roof, or a roof consisting of tilted parts, the maximum deviation is the difference from the mean height to the lowest or highest point respectively. In the following examples the median was chosen.

Depending on their (height-) standard deviation buildings in the image are coloured as shown in Table 4-1 and draped over the final DCM. This way the user gets an impression how reliable the building heights are. In cases buildings are coloured in red the user should proceed to further investigations and check e.g. in the DSM if this designation is reasonable. This is the case for the example in Figure 4-19. In Figure 4-20 two more examples are shown to illustrate the correspondence between DSM quality of matched stereo imagery and the reliability of the chosen height for the buildings in the DCM.

Table 4-1: Colours for height standard deviations ($\sigma_h \approx GSD_{xy}$).

$> 3\sigma$	
$2\sigma \leq n < 3\sigma$	
$1\sigma \leq n < 2\sigma$	
$0 \leq n < 1\sigma$	

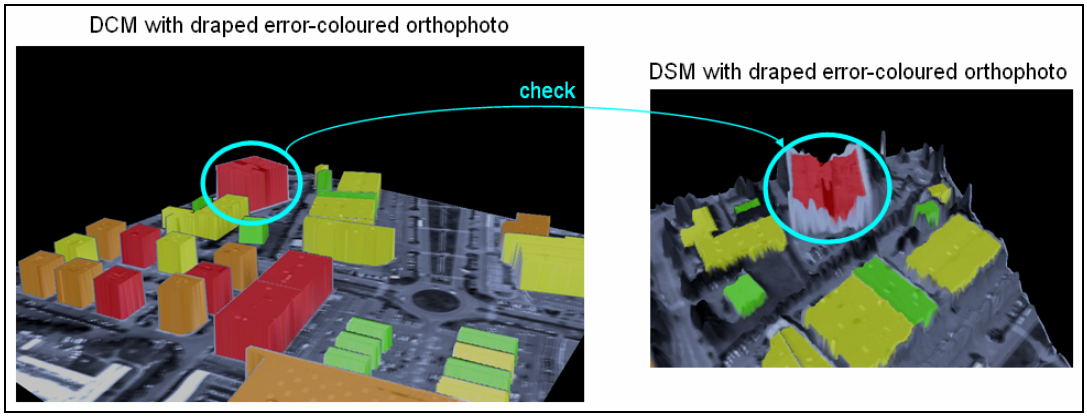

Figure 4-19: Coloured DCM and DSM (ADS40 subset of Nimes).

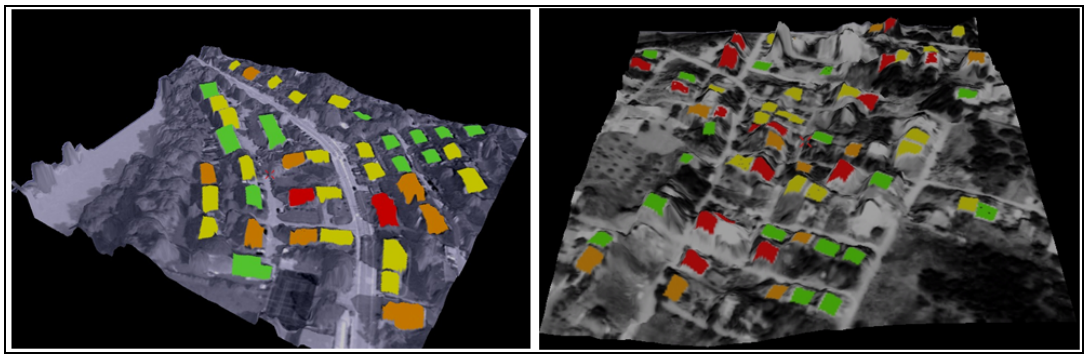

Figure 4-20: Coloured DSMs (left: HRSC subset of Bern, right: IKONOS subset of Athens).

4.d.v Strengths of the Hough Transformation, Examples

4.d.v.A Noisy Data

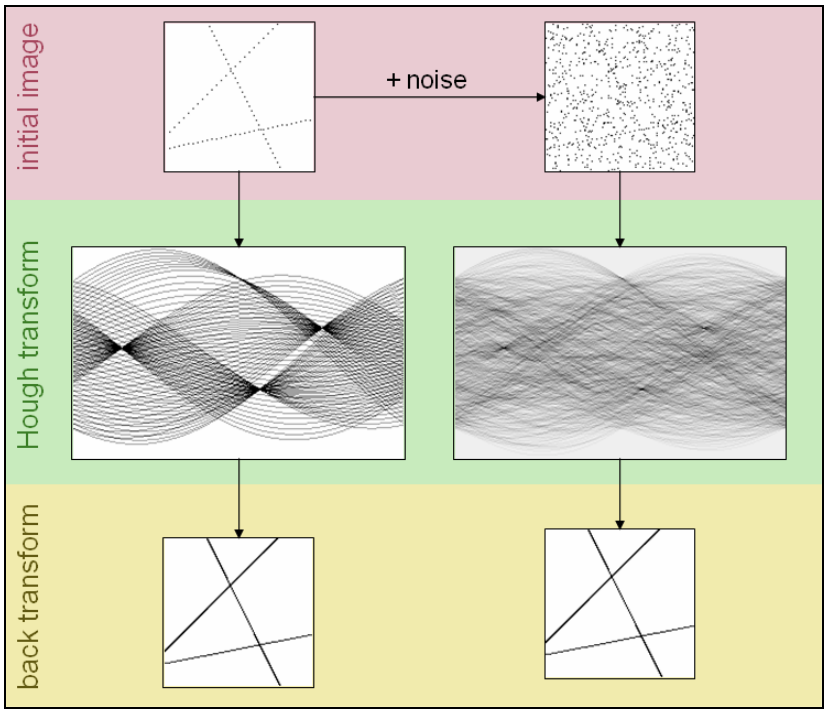

Figure 4-21: Noisy example of Hough and back transform.

The Hough transformation is a very powerful tool, especially when it comes to working with data sets that are noisy [IDL (2004)]. Figure 4-21 illustrates an example, where the original image consists of points lying on three lines. The second initial image is identical but additionally salt-and-pepper (random) noise was added. In this image it s rather difficult for the human interpreter to recognize the original three lines. The next step is a transformation into Hough domain. For each point one sinusoidal curve is drawn. The intersections of the sinusoids characterize points lying

on a common edge (in image space). The higher the amount of such common points on one line is, the more sinusoids intersect at the corresponding location in Hough domain. Then, the local maxima are automatically detected by setting a threshold of 20. This means that it is searched for lines, on which at least 20 points lie. The found local maxima are then back transformed and the final outcome is in both cases identical. The noise was completely disregarded.

4.d.v.B Level of Detail

Another big advantage is the level of detail that can be 'defined' for the final back transformed building. The implementation in TABU gives the user the possibility to choose between two options for defining the level of detail (Figure 4-22).

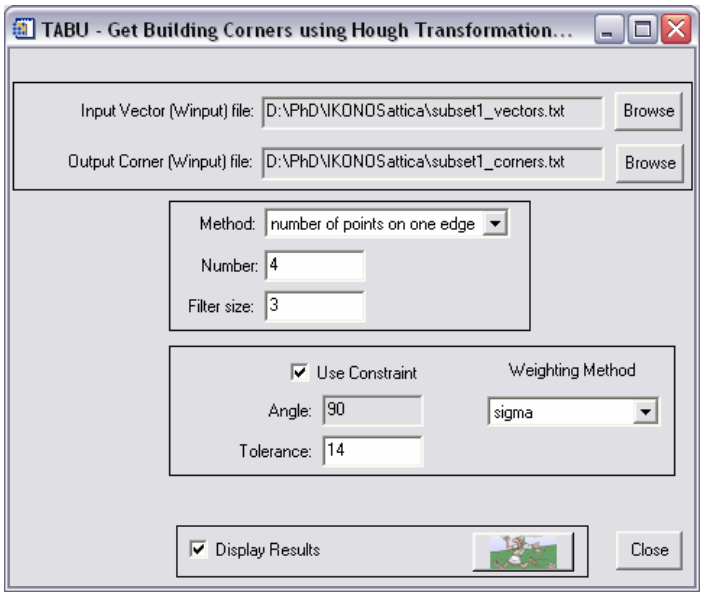

Figure 4-22: TABU-screenshot of the building extraction module.

Users can define either the number of edges to be found, or the minimum number of points lying on one edge. If, for example, the second option is chosen and the number of points is set to 2, the program will search so long for local maxima until no more points are left in input data. Another factor is the chosen filter size that defines the 'elimination'-band width around a found edge. If this filter size is set to 1 (band width is 1 pixel) at each iteration step only a few pixels will be erased, hence more edges and higher detail will be preserved.

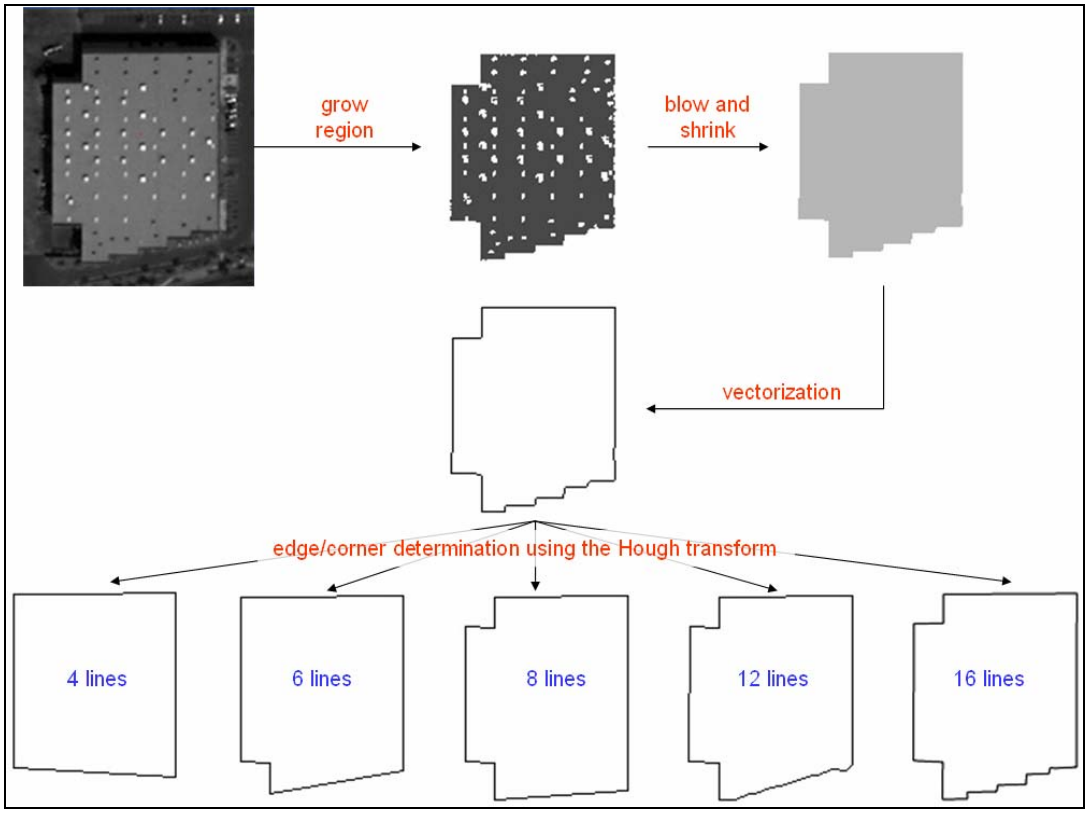

Figure 4-23: Example 1: Level of detail / generalisation possibilities by employing the Hough transform.

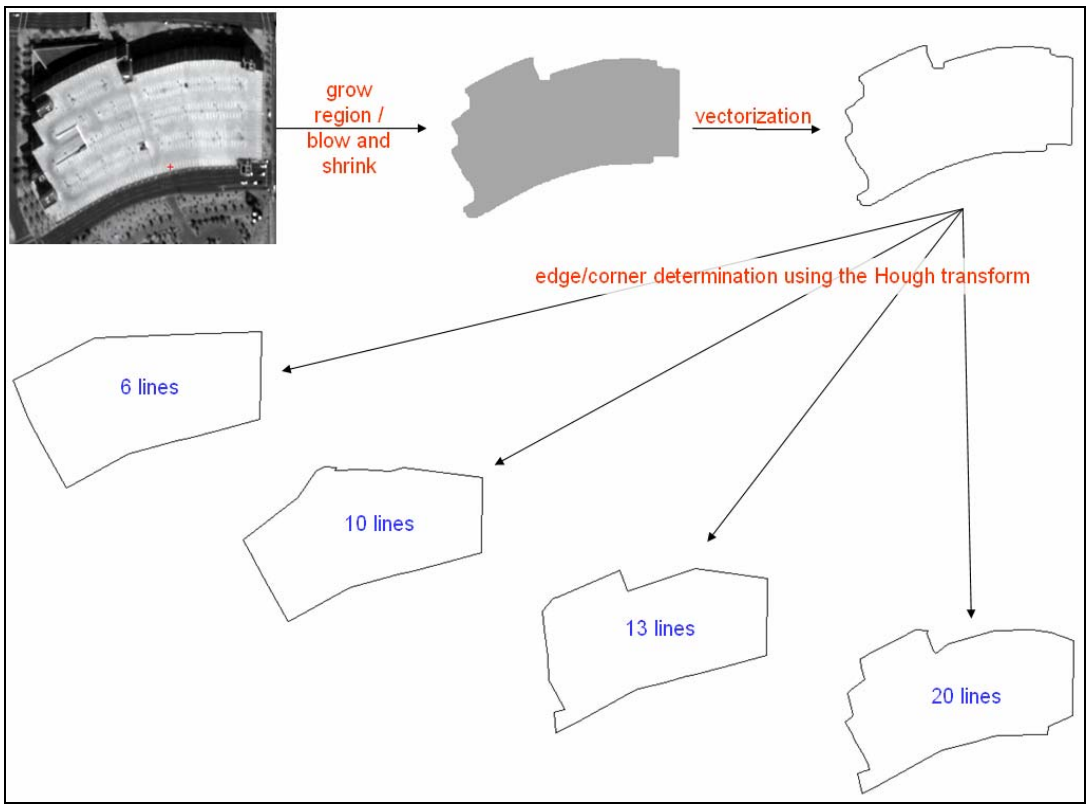

Figure 4-24: Example 2: Level of detail / generalisation possibilities by employing the Hough transform.

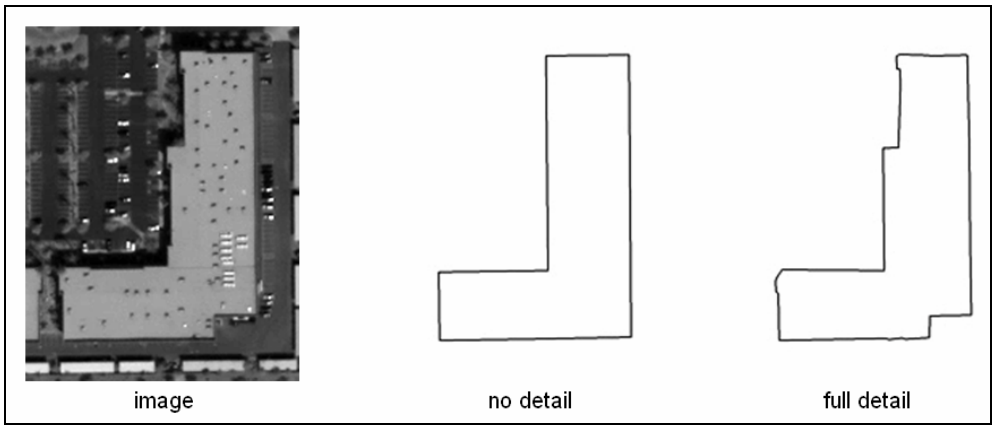

Figure 4-25: Building extracted with no and with full detail.

The presented examples also demonstrate the way in which this approach can be used for generalization purposes.

4.d.v.C Forcing Angles

Quite often buildings are characterized by certain geometric properties. For example, two walls intersect at a certain angle. In the building extraction process it may happen that this geometric property is lost, due to inaccuracies of the input data, inappropriate algorithm strategies or other sources of error. Despite that, we want the extracted building, to obey these characteristics, so it represents reality in the best way.

When using the Hough transform as a tool for creating DCMs this idea can be integrated quite easily. In TABU the user can choose the angle that is requested between intersecting lines and a tolerance value (see Figure 4-22). This tolerance value describes how big the maximum deviation from the requested angle can be, so that it is still forced into the right position.

The implementation is quite easy, since the horizontal axis in parameter space represents angle θ. If the angle between two found edges has to be of a certain value, the horizontal distance between the corresponding local maxima in parameter space has to be equal to the wanted angle. Hence the deriving of the local maxima for the final back projection into image space is adjusted through the "angle" restriction.

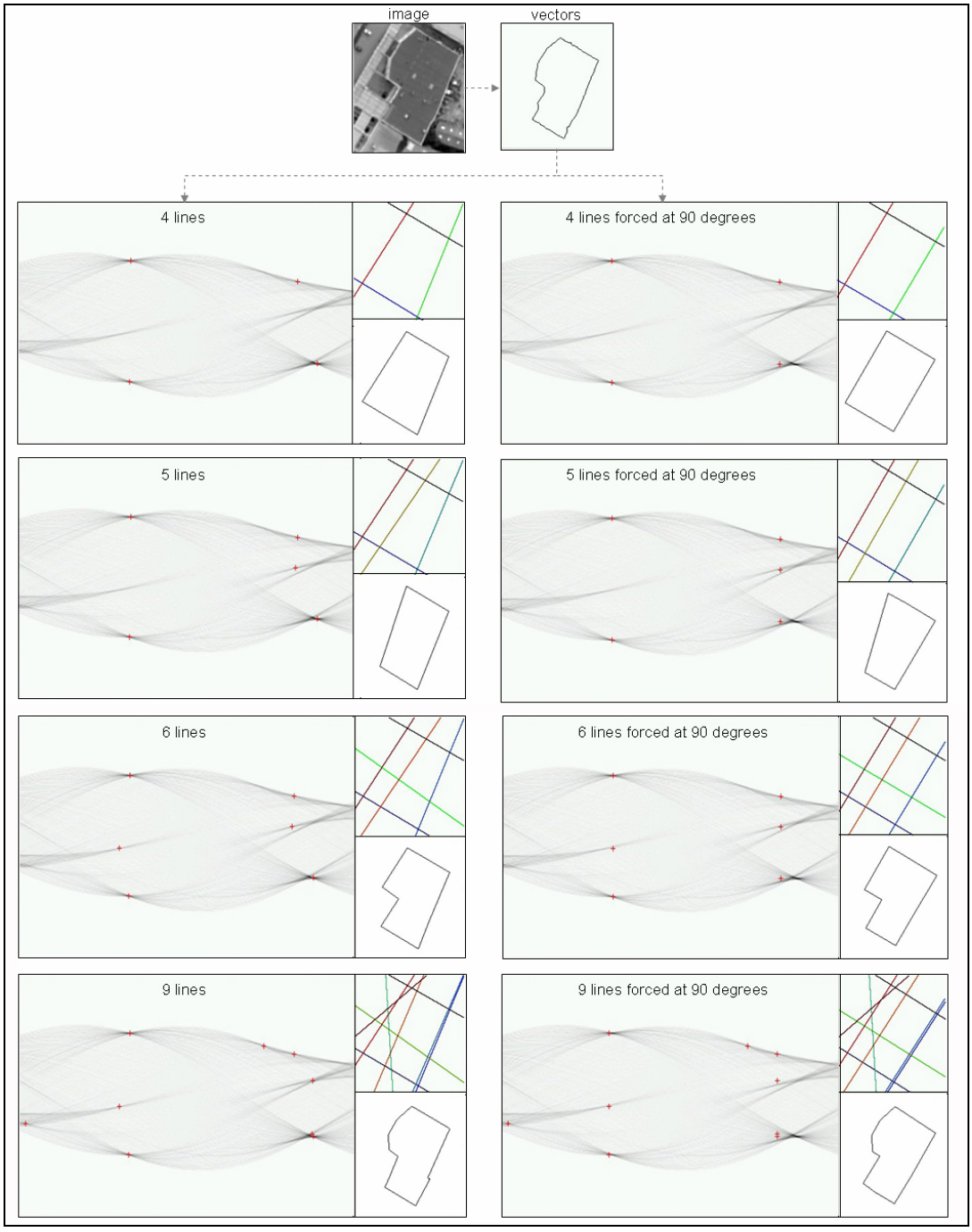

Figure 4-26: Example for applying angle restriction during building extraction.

The final locations of the used points in Hough domain are calculated trough an adjustment. Two solutions to derive the influence (weights) for the positioning of each point are possible:

1. The number of intersections at the local maxima is directly related to the weight.

2. If we want all lines of the building to be either orthogonal or parallel to each other:

   - For each local maximum compute a new $\theta_{i\text{-}new}=(\theta_i - \theta_{min}) \ mod \ \pi/2.$
   - Find the mean θ_{mean} of all $\theta_{i\text{-}new}$.
   - The inverse of $(\theta_{i\text{-}new} - \theta_{mean})$ is the weight given to θ_i in the adjustment.

Figure 4-26 shows an example where a building is extracted under different conditions. The result is depicted on the right side without any constraints and on the left side with a constraint forcing an angle of 90 degrees.

The example in Figure 4-27 shows what happens if the tolerance value is chosen too big. The outcome on the right side of the figure is incorrect. All pairs of neighbouring lines intersect at a right angle which is not correct for the red line. When decreasing the tolerance value (left result) all the line pairs, except the red line, intersect at an angle of 90 degrees.

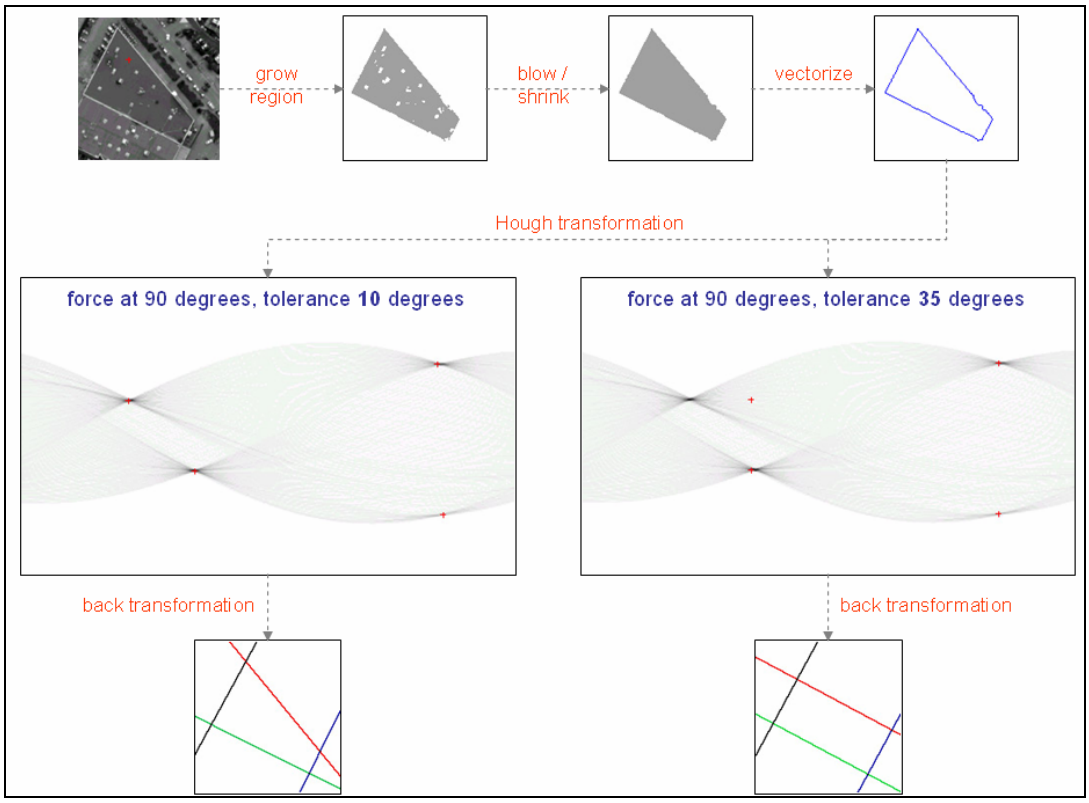

Figure 4-27: Example illustrating erroneous result if tolerance value is chosen too large.

The same way, as we introduced constraints for the horizontal axis in Hough domain, namely in θ-direction, we could also introduce constraints for the vertical (ρ-) direction. This would make sense in cases, where the buildings to be extracted contain parallel edges at a certain distance. During the determination of the local maxima we would have to look that the vertical distance between them (in Hough domain) is equal to the perpendicular distance of the corresponding parallel edges (in image space).

4.d.v.D Bridging Gaps

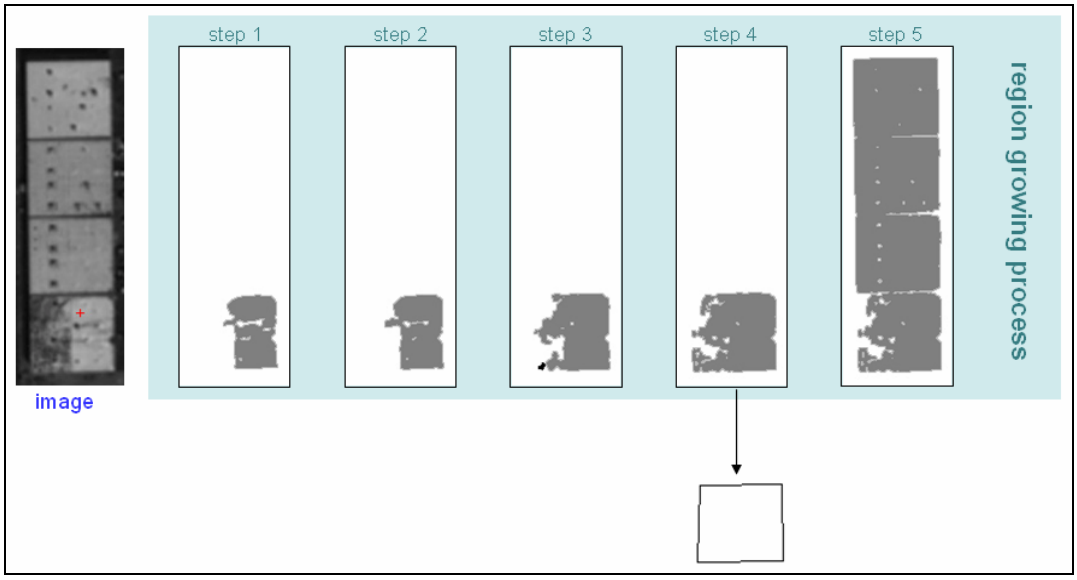

Figure 4-28: Example showing how the algorithm overcomes missing data.

The last advantage that is listed here focuses on missing / incomplete data handling. If the grey value differences inside a building are bigger than the differences between grey values inside and grey values outside the building, the adaptive region growing algorithm may fail to find all relevant pixels of a building. For example, in Figure 4-28 various steps of the region growing process are depicted. The seed point is marked with a red cross in the image. In step 5 the grown region overflows the building edge and a big area is accumulated that does not belong to the investigated building. Hence the region growing process is terminated and the found region of step 4 is used for further processing. It is obvious that the top left corner is not identifiable at all in this region. Even more it is also rather difficult to imagine the left vertical edge. Nevertheless, when applying the proposed building extraction algorithm this edge is found without difficulties. By intersecting it with the upper horizontal edge, the top left corner point of the building (which is not visible at all in the grown region) is reconstructed without effort. This problem is very common, since many times parts of buildings are covered by shadows or vegetation, and hence some of the corner points are not visible.

4.e Change Detection and Updating

"In almost all areas of our society there is an increasing demand for up to date digital map data bases. [...] Map updating can be carried out by complete remapping of the area for each revision cycle, but much work and time can be saved by detecting changes..." [Olsen (2004)]. SPOT (2003) also stresses the importance of rapid map updating, especially in developing countries.
Change detection can be performed on multi-temporal data or by comparing data of a single epoch to an existing map [Murakami et al. (1998)].

Most of the change detection procedures (see *2.d Change Detection and Updating*) aim at identifying changes of existing objects. The drawback is that new objects that do not have any reference in the old data base will not be found. Thus, in these cases we should not speak of "updating", but of "object revision".

Our approach is able to locate new buildings, since in the preceding work a whole new DCM is created. The difficulty we face is to find out which buildings of the old state and new state correspond to each other. Since the new DCM is derived without considering the old state, no reference between corresponding buildings in the new or old state exists yet.

General approaches to detect corresponding buildings in different states are:

- Compute the centre of gravity of each object in both data bases and compute the Euclidian distances between "gravity-centre-points" of both states: distances shorter than a certain threshold correspond to buildings that belong to each other.
- Calculate the distance between centroids of the bounding ellipses [Post (1984)].
- Take the building edges of the old state and find the closest lying building outline of the new state.
- Check the overlapping areas of each building between both states. If this overlap exceeds a certain limit the buildings are considered to be the same.

For our change detection and updating strategy we employ the latter solution and suggest an overlap threshold of 60% (see Figure 4-29). As soon as all buildings in the data set have been checked, one can easily say whether buildings have disappeared or if new ones were built.

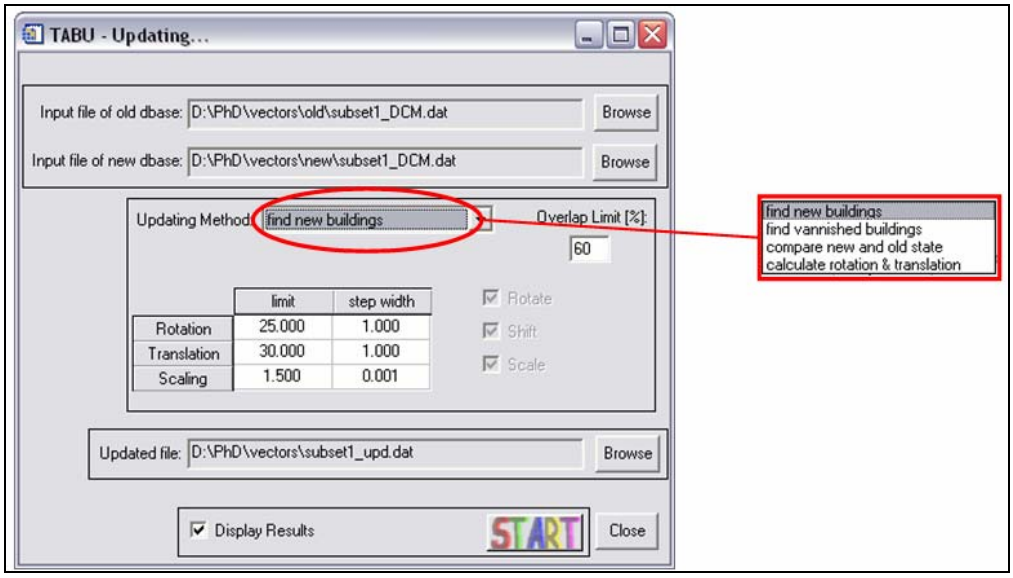

Figure 4-29: Graphic user interface of the implemented Change Detection and Updating module.

For all found buildings that exist in both the new and the old state and additionally correspond to each other a comparison calculation is performed.

In TABU users have four choices (see Figure 4-29):

- Detect new buildings,
- Detect vanished (=demolished) buildings,
- Compare new and old state through a residual computation between the building corners, or
- Compare new and old state by calculating the rotation, translation and scaling parameters that describe the transformation from the old state into the new state of the building.

It may occur that the number of building corners of a single object may differ in the old and new state. If this is the case, the residuals are calculated from the building having less corner points, to the closest lying corner-point of the corresponding object. Each corner point can be used only once for the residual calculation.

The previous step, the overlap calculation, gives an indication how high the correspondence between buildings of different states is. Even if the overlap is very high it does not mean necessarily that the building of the old state still exists in the new state.
Figure 4-30 shows some examples of corresponding buildings. All these buildings have "passed" the overlap-test. Examples 1,2,3 and 4 show a typical behaviour, were

reasonable residuals are calculated. Contrarily, example 5 shows a case, were the new and old building do not correspond.
This case can be detected by analyzing the corner residuals. There are two suggested approaches: Either analyse the residual lengths independently or look at the sum of residuals of each building. Possible automated determinations of scene dependent thresholds defining the acceptance of a building are given below:

- Calculate the mean of all residuals in the scene and the residual standard deviation. Then check the deviation of each residual from the mean. If it is greater than twice the standard deviation, mark the building as "not" corresponding.
- Compute the sum of residuals for each building. Get the mean of these sums. If the residual-sum of a building differs more than twice the standard deviation from the mean, mark the building as "not" corresponding.

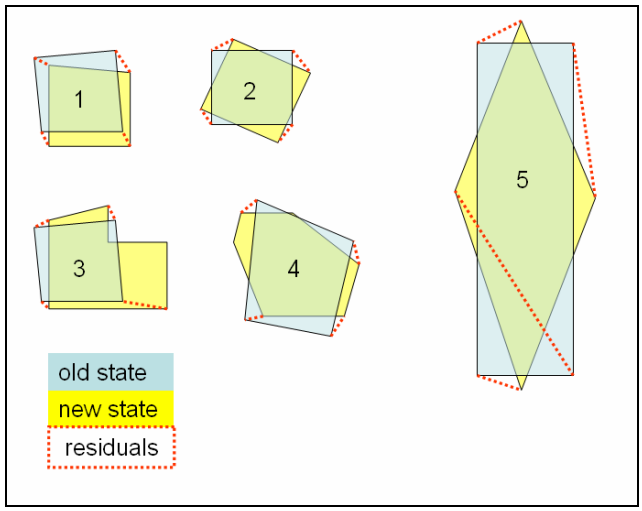

Figure 4-30: Residuals between old and new buildings.

Available 3D information also allows us to make statements regarding changes of building heights. Hence e.g. houses with newly built storeys can be detected, although the building outline remains identical.

This updating module can also be employed for quality control purposes.

5 Alternative Approaches

There is no perfect algorithm for creating DCMs and, therefore, it is advisable to provide (possibly a set of) alternative approaches or even generate a rule-based system which is able to select the most suited approaches or an optimal combination of them for a certain type of residential regions. This chapter is dedicated to investigating two alternatives for cases where the previously proposed strategy fails.

Usually the quality of the input data (i.e. imagery and DSM) is responsible for erroneous results. It is very difficult for the introduced algorithm to produce correct results, if the building to be extracted does not cover a certain number of pixels. For instance, when dealing with IKONOS imagery (with their GSD of 1m) a small house of 8m x 10m will most probably not be extracted correctly. Moreover, the accuracy and resolution of the derived DSM - and nDSM - from such a data set will make it difficult to find adequate seed points for building determination.

The procedures discussed in this chapter fall into the category of feature extraction and make use of the imagery for deriving the geometric building properties.
The general idea is to use a library where characteristic features of buildings are stored; by analyzing the image, areas are searched that correspond to a high degree to the registered "library buildings". Characteristic features of a building can be textural measures (by using the so-called occurrence and co-occurrence descriptors) or similarity measures of the image grey values.

Two techniques are discussed; the first one is based on a matching strategy and the second technique is based on pattern recognition.

5.a Building Extraction by Image Matching

The strategy follows the basic principle of image matching by correlation. A given reference image matrix is searched in the image under investigation (the so-called search image) by moving the reference matrix pixel by pixel over the entire image area. Potential candidate positions, i.e. positions of high similarity, are marked if a so-called correlation coefficient exceeds a predefined threshold. In order to find the optimum geometric fit, the searching procedure includes, besides translation, also rotation. It would also be possible to include scaling, meaning that houses of similar shape but different size are searched too.

The reference image is usually a small image matrix, here depending on the size of the building to be searched, whereas the search image is a rather big image matrix in our case covering the whole area under investigation.

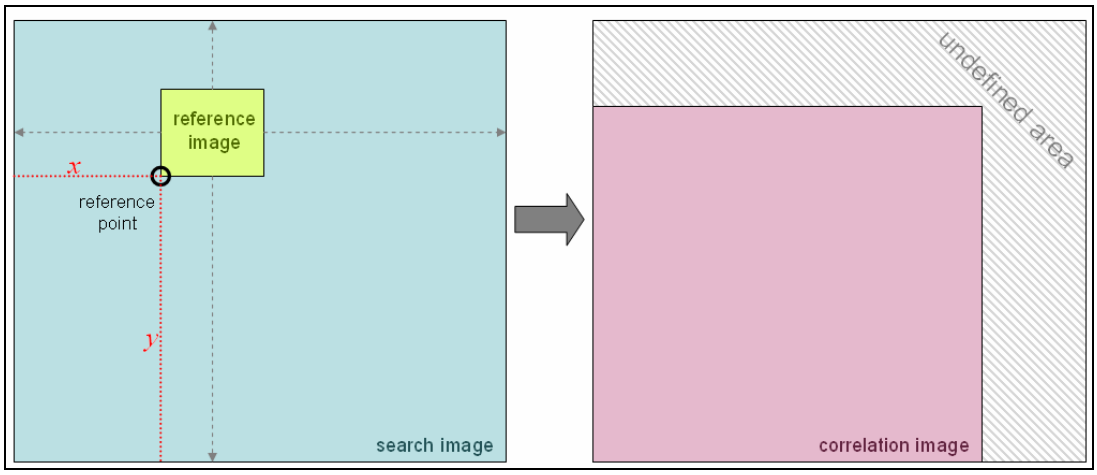

Figure 5-1: Search, reference and correlation image.

Figure 5-1 shows the principle of the correlation procedure. The left hand side indicates the searching process with the reference image and the given spaceborne or airborne image as search image. The correlation index is computed for each position of the reference image and the results are stored as similarity measure in the so-called correlation image. Potential building positions are characterized by a high correlation coefficient and thus the correlation image just needs to be thresholded and the local maxima are localised. It has to be mentioned that one crucial parameter is certainly the appropriate threshold value. Its choice determines quite significantly the quality of the result. If the threshold is too low, too many buildings are detected leading to a great number of false matches. If the threshold is too high, the selection is too strict and, as a consequence, too many buildings will be rejected. It is not possible to define an optimum threshold as a general suggestion.

For the cross-correlation coefficient using 0.7 to 0.8 is certainly a good choice for starting, but individual adjustments are necessary in any case.

As measure of similarity the cross-correlation coefficient (Equation 5-1) is adopted, but also other measures can be used (Equations 5-2 and 5-3) [Kraus (1990)].

$$c_1 = \frac{\sum (g_1 - \bar{g}_1) \cdot (g_2 - \bar{g}_2)}{\sqrt{\sum (g_1 - \bar{g}_1) \cdot \sum (g_2 - \bar{g}_2)}} \tag{5-1}$$

$$c_2 = \frac{\sum \left(|g_1 - g_2| / (g_1 + g_2) \right)}{n} \tag{5-2}$$

$$c_3 = \sqrt{\frac{\sum (g_1 - g_2)^2}{n}} \tag{5-3}$$

where g_1 and g_2 are the grey values in the reference and search window,

$\bar{g}_1$ and $\bar{g}_2$ are the mean grey values in the reference and the search window and

n is the number of used pixels.

Kraus (1996) suggests rewriting Equation 5-1 as follows for a more efficient computation:

$$c_1 = \frac{\sum g_1 \cdot g_2 - n \cdot \bar{g}_1 \cdot \bar{g}_2}{\sqrt{\sum \left(g_1^2 - n \cdot \bar{g}_1^2 \right) \cdot \sum \left(g_2^2 - n \cdot \bar{g}_2^2 \right)}} \tag{5-4}$$

Note, that when using Equation 5-4 the computing effort is reduced since the expression $\sum \left(g_1^2 - n \cdot \bar{g}_1^2 \right)$ is constant during the whole process and has to be calculated only once.

The finding of the maxima in the correlation image with sub-pixel accuracy by approximating the discrete correlation function by a continuous polynomial function is broadly discussed in Kraus (1996) and will not be described here.

Figure 5-2 shows an example were one search image was matched with multiple reference images.

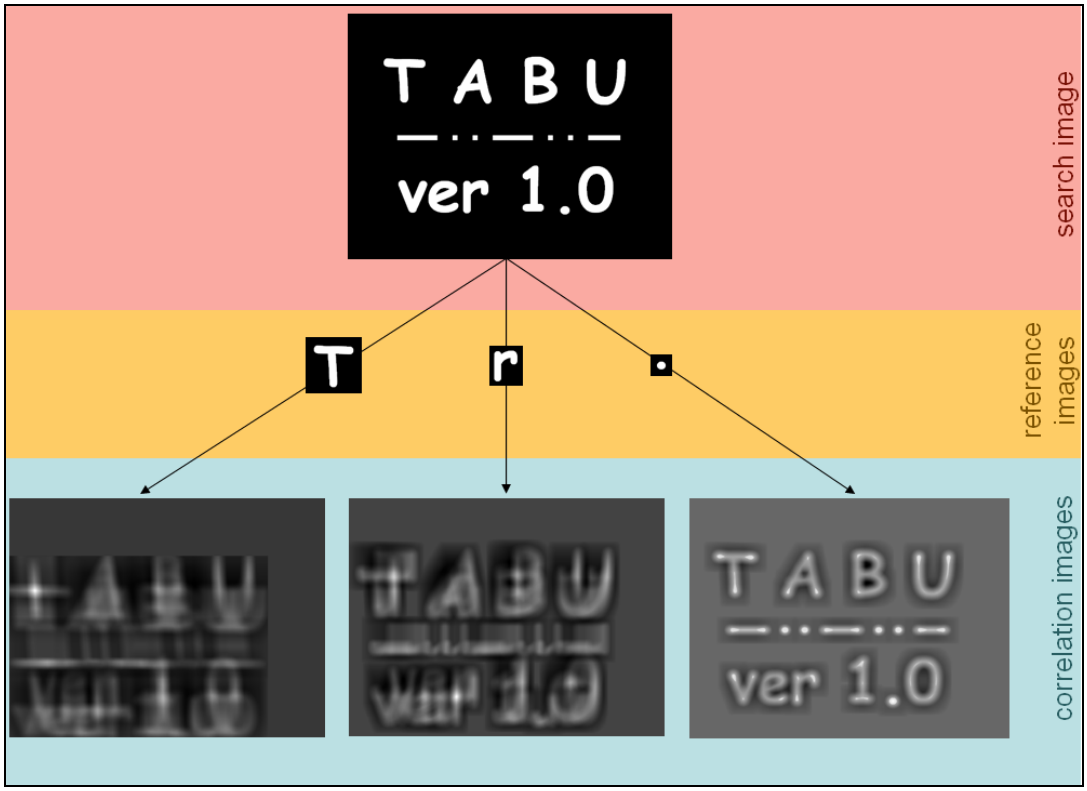

Figure 5-2: Correlation example. The brighter the pixels, the higher the correlation, i.e. the greater the similarity between search image and reference image.

As far as building extraction is concerned, the reference images of buildings have been stored in a library. They can either be extracted from the current scene or might be taken from an external source. For each of these buildings also vector information (describing the building outline in the reference image coordinate system) is either digitized on the reference image or imported if available.
At this stage the library contains multiple building types to which one image patch and one vector representation corresponds.

Once a location of high correlation is found in the search image, the vector data of this building is transformed into the coordinate system of the search image.

It is often the case that buildings of the same (or similar) shape have different colours (grey values) in the images (e.g. due to different roof materials). Therefore it is advisable not to store image patches of the investigated buildings in the library, but instead, register their edges. In this case, also the search image has to be edge-extracted before applying the matching procedure. For gaining the edge information, classical operators like the Canny edge detector, Sobel operator, Laplacian of Gaussian etc. can be applied.

The implementation in TABU (Figure 5-3) is straight forward. The user defines a search image and a patch (=reference image) (either by digitizing it on the search image or by selecting a file) that will be used for the matching. Moreover, the user chooses in what intervals and/or how many rotations shall be performed while moving the patch over the search window. For symmetric buildings the total rotation should be 180 degrees whereas for unsymmetrical buildings 360 degrees. This is necessary since we do not want to store multiple reference images of the same building with all possible orientations in the library; thus the program rotates the reference patch by a certain degree when moving over the search image. Obviously, the bigger the number of rotation steps and the smaller the step widths become, the higher will the detection and output accuracy be.

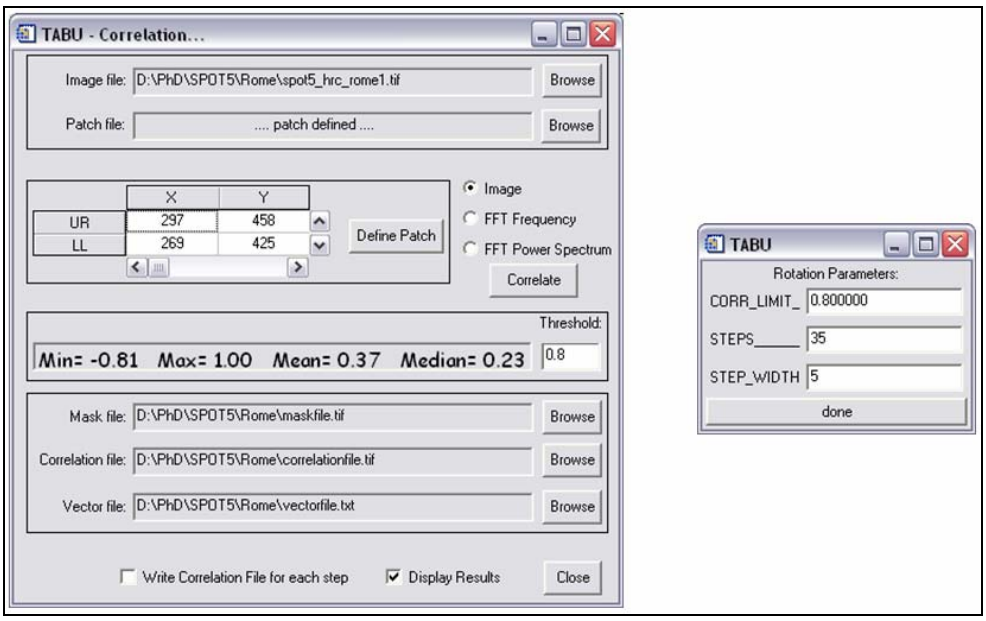

Figure 5-3: Screenshot of the TABU interface for the correlation.

In this case (Figure 5-3) the moving reference window will be rotated 35 times by 5 degrees at each position in the search window. In case of a building with a rectangular outline this should be sufficient. Hence we obtain a correlation image with 35 layers, each holding the correlation coefficients corresponding to a certain orientation of the moving patch. Then, in this multi-dimensional correlation file it is searched for all values exceeding a certain threshold (in this case 0.8), and the positions are marked as found buildings.

The vector information (in the search image coordinate system) for each building is used for building up the DCM. The height value for each building is taken from the corresponding nDSM.

Figure 5-4 and Figure 5-5 show two examples of the image matching procedure.

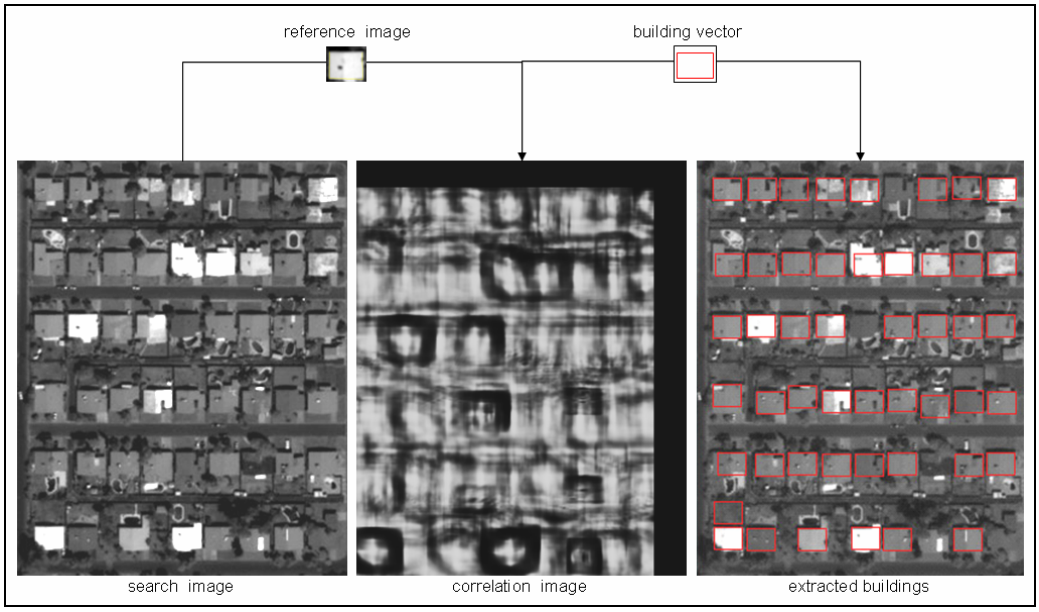

Figure 5-4: Example on image matching with no rotation (Quickbird-subset of San Diego).

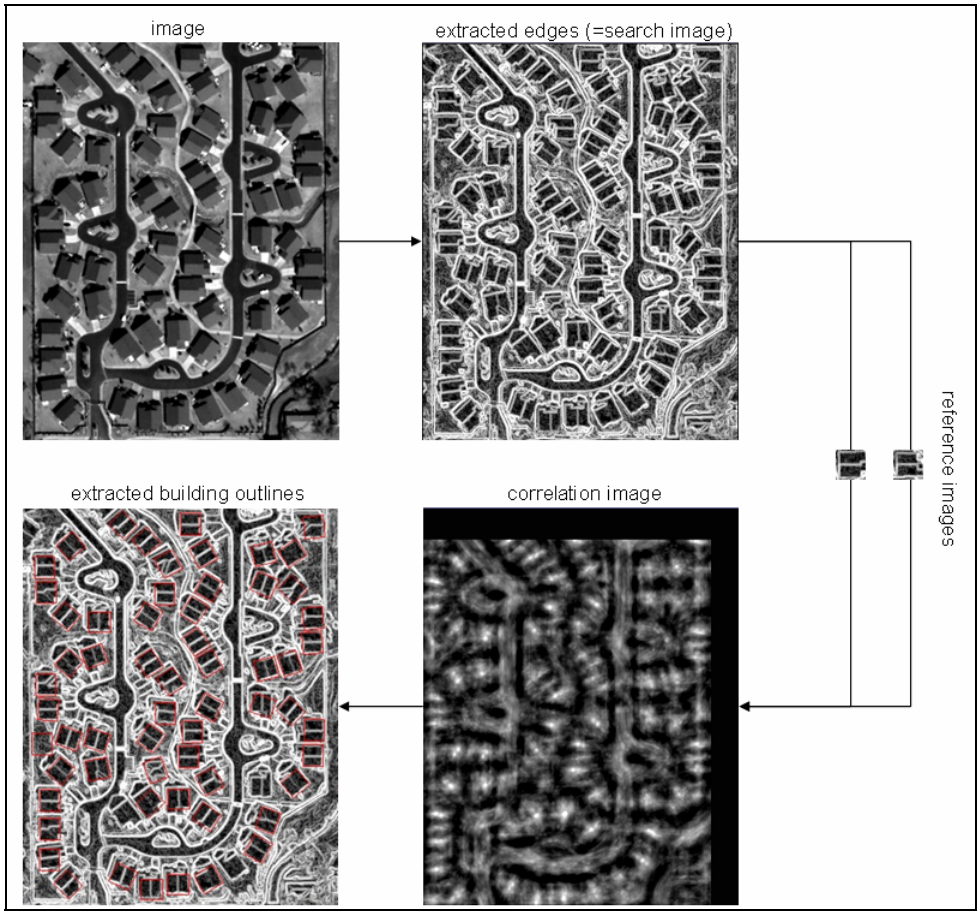

Figure 5-5: Example on image matching with two reference images and multiple orientations (Quickbird-subset of Phoenix area). Note that the computed correlation image is a multi-dimensional image. The number of dimension corresponds to the number of rotations (here: 119 with a rotation step of 3 degrees).

5.b Building Extraction with Texture Analysis

One of the simplest ways for describing texture is to use statistical moments of grey level histograms of an image or a region. Measures of texture computed using only histograms suffer from the limitation that they carry no information regarding the relative position of the pixels with respect to each other. One way to bring this type of information into texture analysis process is to consider not only the distribution of intensities, but also the distribution of intensity variation [Gonzalez and Woods (2002)].

For this kind of textural examination, firstly the so-called co-occurrence matrix has to be derived for the examined area. This particular matrix holds e.g. information of pixel changes in multiple directions (usually horizontally, vertically and diagonally). The co-occurrence matrix' extents are same in both directions and equal to the number of grey levels that will be considered. For example, for an 8 bit image (256 grey values) the co-occurrence matrix' extents would be 256 by 256. Usually a recoding is carried out to reduce the number of grey value classes (also called bins). A recoding of the original image down to 16 grey levels is for most of the cases satisfying [Gong et al. (1992)]. Nevertheless, during this research (on texture analysis), all images were recoded to 40 bins.

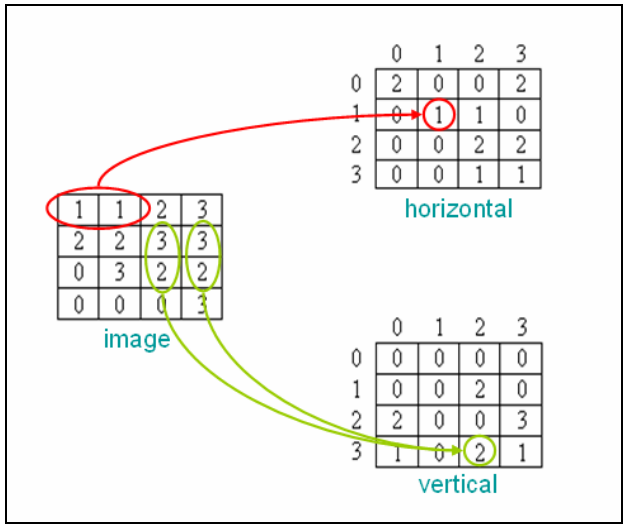

Figure 5-6: Image and corresponding co-occurrence matrices in horizontal (left to right) and vertical (top to down) directions.

At each position $m_{c,r}$ the co-occurrence matrix holds the number of changes from class r (=row indices) to class c (=column indices) (see Figure 5-6). This computation is carried out for multiple directions, meaning that one co-occurrence matrix is created for each direction. The following example illustrates the creation of

such matrices; here, four grey values exist and the co-occurrence matrices are derived for horizontal and vertical directions.

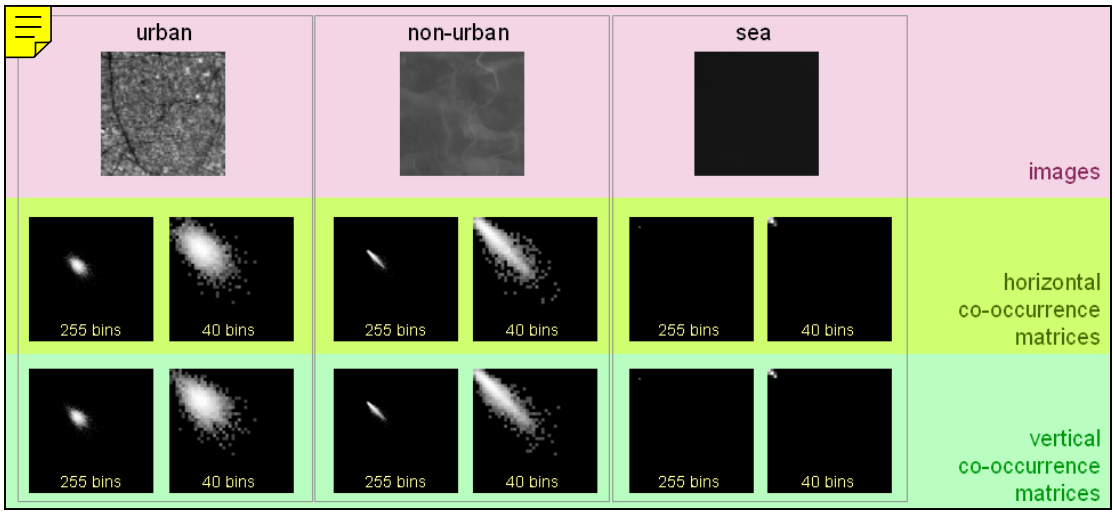

Figure 5-7: Horizontal and vertical co-occurrence matrices for 3 different types of terrain.

The task now is to analyze a given co-occurrence matrix in order to categorize the region for which it was computed. Therefore descriptors are needed that characterize these matrices. It is obvious how important it is to include some kind of information that tells us whether the values are well distributed over the whole matrix, or whether they are mostly located close to the matrix diagonals (e.g. Difference Moment or Inverse Difference Moment, Equations 5-10 and 5-11). In case big values lie close or on the diagonal (Figure 5-7, non-urban), the region under investigation is expected to be homogeneous, whereas if the values are distributed more homogeneously (Figure 5-7, urban), the co-occurrence matrix corresponds to a heterogeneous region.

Some of the most commonly used descriptors are [Haralick (1979), Gonzalez and Woods (2002), Zhang (2001)]:

$$Maximum\ Probability = \max_{I-1,J-1}\left(f(i,j)\right) \tag{5-5}$$

$$Contrast = \sum_{i=0}^{I-1}\sum_{j=0}^{J-1}(i-j)^2 \cdot f(i,j) \tag{5-6}$$

$$Uniformity = Angular\ Second\ Moment = Energy = \sum_{i=0}^{I-1}\sum_{j=0}^{J-1} f(i,j)^2 \tag{5-7}$$

$$Entropy = -\sum_{i=0}^{I-1}\sum_{j=0}^{J-1} f(i,j)\cdot \log_2\left(f(i,j)\right) \tag{5-8}$$

$$Homogeneity = \sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \frac{f(i,j)}{1+|i-j|} \tag{5-9}$$

$$Difference\ Moment\ of\ order\ k = \sum_{i=0}^{I-1}\sum_{j=0}^{J-1} (i-j)^k \cdot f(i,j) \tag{5-10}$$

$$Inverse\ Difference\ Moment\ of\ order\ k = \sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \frac{f(i,j)}{(i-j)^k} \qquad i \neq j \tag{5-11}$$

$$Mean = \overline{f(i,j)} = \frac{1}{I\cdot J}\cdot \sum_{i=0}^{I-1}\sum_{j=0}^{J-1} f(i,j) \tag{5-12}$$

$$Variance = \frac{1}{(I\cdot J)-1}\cdot \sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \left[f(i,j)-\overline{f(i,j)}\right]^2 \tag{5-13}$$

$$Skewness = \frac{1}{I\cdot J}\cdot \sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \left(\frac{f(i,j)-\overline{f(i,j)}}{\sqrt{Variance}}\right)^3 \tag{5-14}$$

$$Kurtosis = \frac{1}{I\cdot J}\cdot \sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \left(\frac{f(i,j)-\overline{f(i,j)}}{\sqrt{Variance}}\right)^4 - 3 \tag{5-15}$$

$$Mean\ Absolute\ Deviation = \frac{1}{I\cdot J}\cdot \sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \left|f(i,j)-\overline{f(i,j)}\right| \tag{5-16}$$

where i is the column and

j the row position in the examined array.

These measures can be applied both on the image subsets themselves in form of a statistical histogram analysis or on the corresponding co-occurrence matrices.

The approach for building extraction is similar to the one explained for the matching process. In a library signatures of texture for different buildings, each with multiple orientations, are stored.

Then a moving window (a so-called kernel) is placed on the image and the textural measures (via co-occurrence matrices) are computed by using the corresponding image values. The texture measures are stored at the position of the reference point in a new image matrix (see Figure 5-1). Then the kernel is moved by one pixel and the procedure is repeated until the whole search image is covered. The result is a "texture" image that holds textural measures in various layers; one layer for each descriptor.

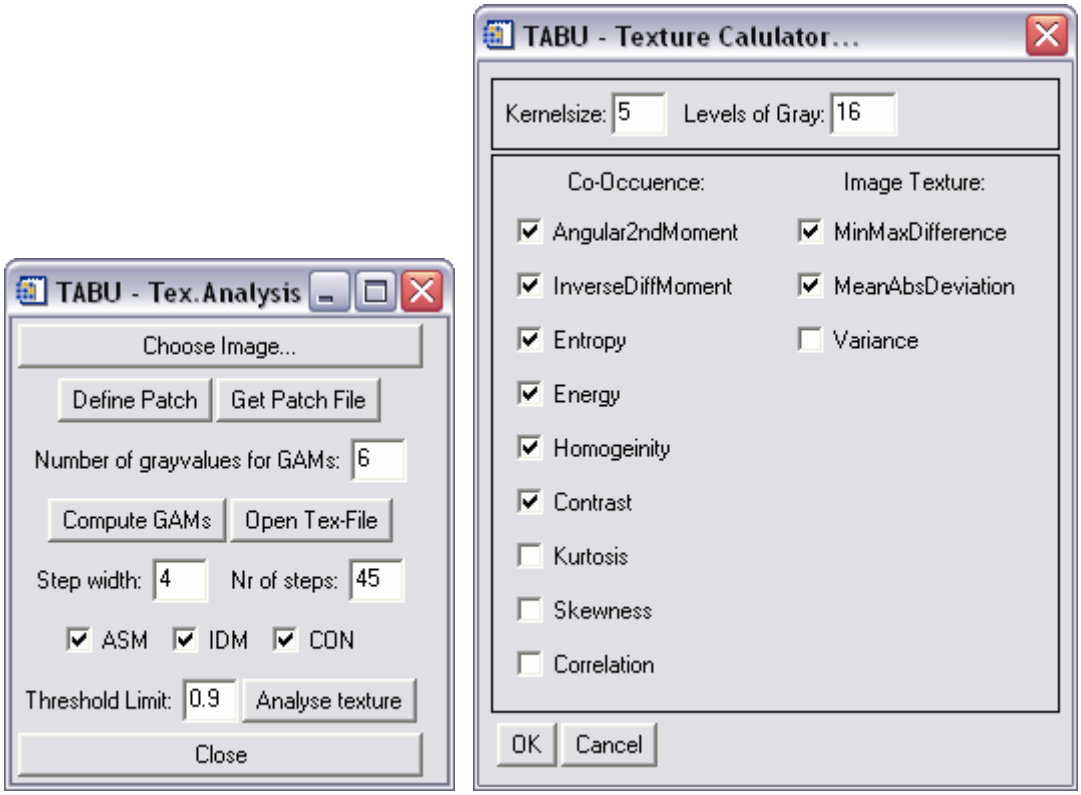

Figure 5-8: Screenshot of the two modules implemented in TABU for texture analysis.

For detecting buildings in the image, the multi-layer "texture" image information is compared to the texture measures stored for a specific building type in the library; this can be done in form of a decision tree. Another approach would be to apply a multi-"spectral" classification of the "texture" image, where training information is taken from textural descriptors of a building in the library.

Both of these approaches are implemented in TABU (Figure 5-8).

The general idea when applying texture analysis was to reduce the processing time if compared to the previously described matching technique. If it is searched for a building in an image, and the building is also rotated at each position, the sum of correlation computations equals to: *number_of_positions* x *number_of_rotations* (Figure 5-9, left).

One can imagine that if the search image is very large and a certain building is searched at 360 different orientations (with a rotation step of 1) the computing effort will become enormous.

When using the texture-analysis approach, the rotations are performed only once for the building to be searched. For each orientation the statistical measures are derived and stored as attributes of the building. Then, the whole search image is investigated: at each position the textural characteristics are computed and compared to the ones stored. If the similarity is big, one is in position to say whether a building is found and should this occasion arrive, what its orientation is. The computing effort would be: *number_of_rotations* + *number_of_positions* (Figure 5-9, right). But be aware that the effort of one processing step in the latter approach is higher since the co-occurrence matrix has to be derived at each position.

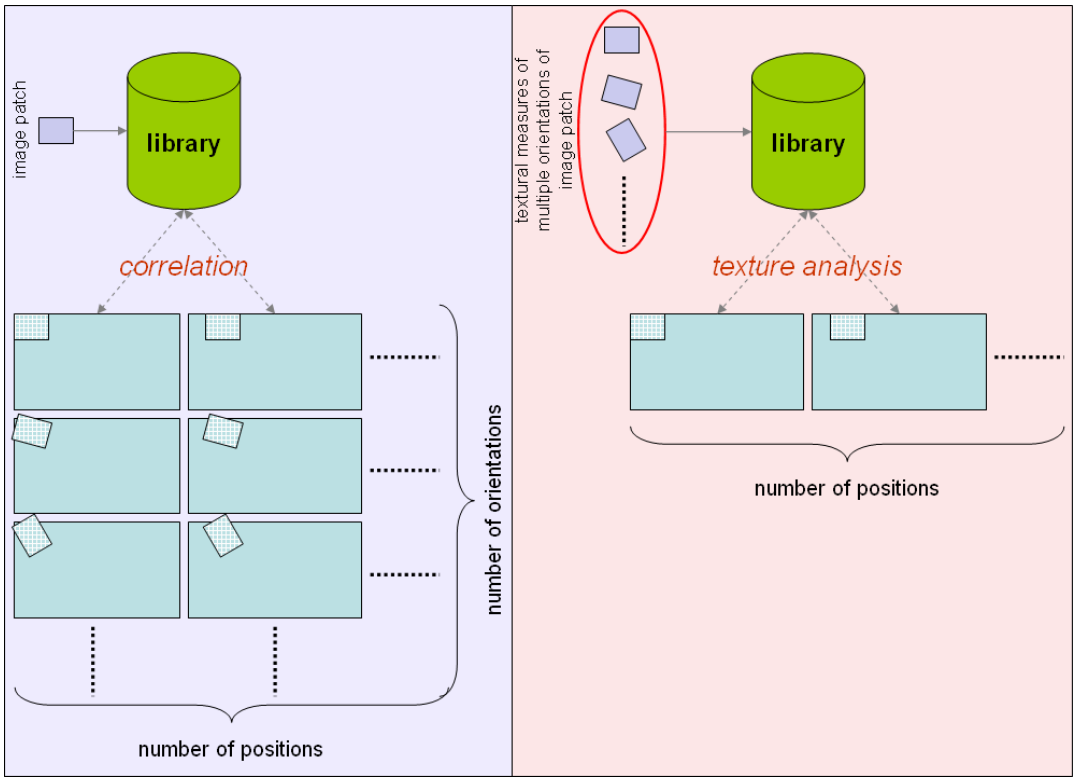

Figure 5-9: Comparison of processing effort (image matching vs. texture analysis).

Tests show that our texture analysis approach is not convenient for extracting individual buildings. The problem is the fact that co-occurrence matrices of different buildings do not differ significantly, so that they can be clearly distinguished through the textural descriptors. Especially when it comes to recognizing a building in different orientation, this method will most probably fail. The reason is that the grey value differences of neighbouring pixels in a certain direction of a building in multiple orientations are very similar.

6 Results and Quality Assessment

In this chapter an evaluation of the presented methods is given. It consists of a quantitative and qualitative description, and moreover shortcomings and weaknesses of the presented methods are discussed.

Figure 6-1 and Figure 6-2 show the input and outcome of the automated building extraction methodology. In the first figure an orthoimage was draped on the nDSM. In the latter one the image was draped onto the derived DCM. It is obvious that each building is associated with one building height (tilted roofs are neglected). Hence for each building its corners and one height attribute are stored in the DCM data base. This information can be easily used for further processing in various research fields.

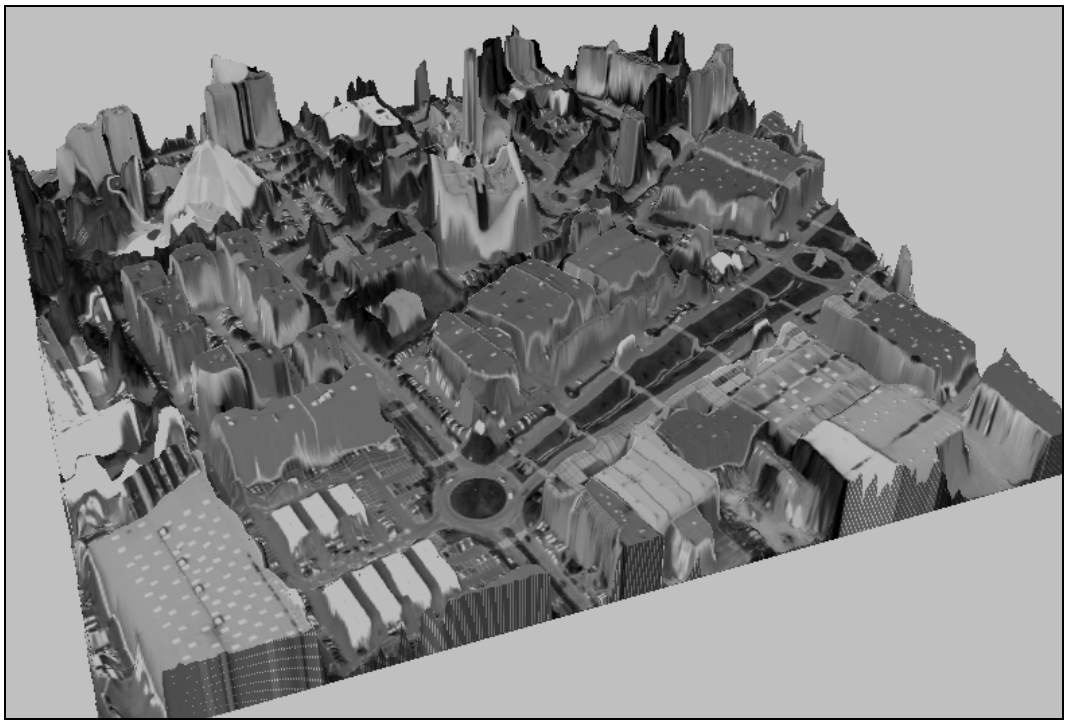

Figure 6-1: nDSM with draped orthophoto (ADS40-subset of Nimes area).

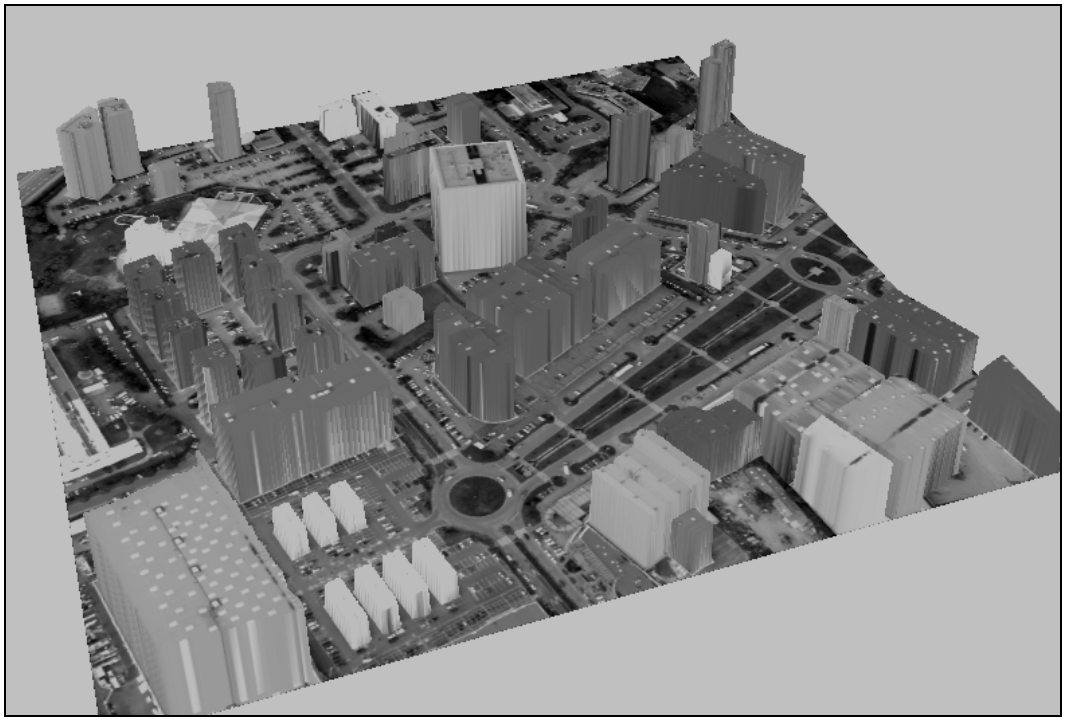

Figure 6-2: DCM with draped orthophoto (ADS40-subset of Nimes area).

Since many subsets are examined that are coming from various types of line scanning systems, both airborne and spaceborne (ADS40, HRSC-AX, Quickbird, Orbview, IKONOS, SPOT5), their outcomes will not be listed individually. Errors in the qualitative evaluation will be given in image space units (pixels).

We consider that the pre-processing has been carried out without error, so that the orientation of the imagery and the derived orthophotos are correct. We will also not evaluate nDSM extraction algorithms and their qualities in detail, since this is not topic of this research.

The presented outcomes are divided into two groups: quantitative and qualitative results. Moreover, the three presented DCM extraction approaches are evaluated individually. Input data is subdivided into categories depending on image scale and building density (low urban and urban) of the investigated areas. Image scale is defined as the scale that we would expect from an analogue product, e.g. for a 1:10,000 product we expect 1-2 metres accuracy in nature, if the graphical accuracy and visual perceptivity are 0.1-0.2mm.

Regarding the mentioned image scales the three interpretation categories are:

1. scale A: 1:1000-1:4000
2. scale B: 1:4000-1:12000
3. scale C: < 1:12000

6.a Quantitative Assessment of Building Extraction

The aim in the quantitative analysis is to evaluate whether the presented approaches are practical in sense of completeness of building detection of the result, i.e. how many buildings were actually found. It is investigated whether the techniques for finding potential building candidates are applicable. Furthermore, an evaluation is carried out to see how many of these buildings were extracted and to what a degree:

- CFB: **C**orrectly **F**ound **B**uildings,
- NFB: **N**ot **F**ound **B**uildings (also includes insufficiently mapped buildings: building seed point was determined successfully, but the adaptive region growing process did not manage to create an area that covers a reasonable amount of the object),
- WFB: **W**rongly **F**ound **B**uildings, i.e. found objects were in reality no building exists.

The calculation of CFB, NFB and WFB are briefly explained in the following:
The CFB and NFB percentages are calculated with respect to the total number of existing buildings in the area under investigation, whereas the WFB is calculated with respect to the total number of found buildings (comprising correctly and wrongly found buildings). Figure 6-3 and Figure 6-4 show the way of computing and a numerical example, respectively.

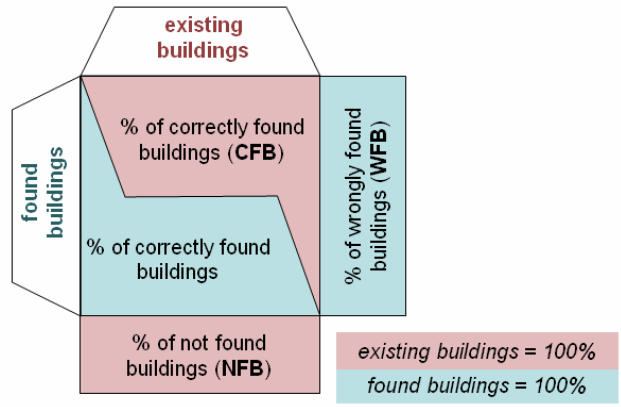

Figure 6-3: Illustration for quantitative assessment computation.

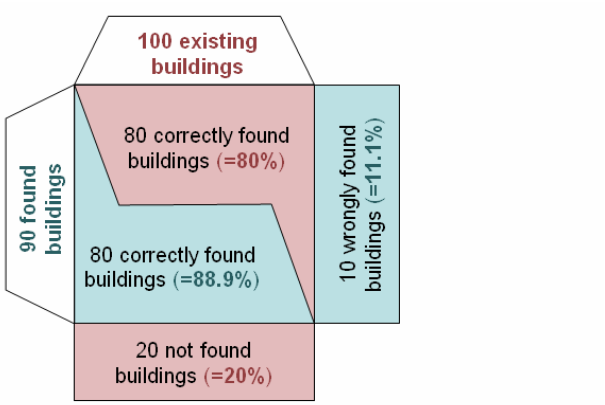

Figure 6-4: Example: 100 existing, 90 found, 80 correctly found, 20 not found, 10 wrongly found buildings. Note the different percentages for the correctly found buildings.

For the evaluation of the outcomes altogether 13 different scenes containing 677 buildings were examined.
Table 6-1 allocates the quantitative analysis.

Table 6-1: Quantitative results.

scale	A	B	B	C	C
type	urban	low urban	low urban	low urban	low urban
sensors	airborne	airborne	spaceborne	spaceborne	spaceborne
method	Seeding	Seeding	Seeding	Matching	Texture Analysis
CFB	80.5%	90.2%	97%	88.3%	34.4%
NFB	19.5%	9.8%	3%	11.7%	65.6%
WFB	7.3%	4.8%	30.3%	1.7%	17.2%

Concerning the level of detail that can be derived from individual data sets the Nyquist theorem has to be taken into consideration ("Sampling rate must be at least twice as high as the highest frequency of the signal."). For our situation it means that sensor's geometric resolution determines the object resolution, or in other words, the level of detail of the object.

6.b Qualitative Assessment of Building Extraction

The qualitative analysis is based on a comparison between the building outlines derived by using the proposed automated methodology, and manually mapped buildings. The manual mapping is carried out by a professional operator who performs 2D (or 3D) digitization on the input data (oriented imagery or orthophotos).

The residuals of each building corner from the manual mapping and the closest point of the automatically extracted shape are computed as a quality measure.

The results are categorized in three groups depending on the method used for building extraction (Table 6-2). The RMS is given in pixels.

Table 6-2: Qualitative results.

	Hough	Matching	Texture Analysis
RMS x	**0.937**	**0.898**	**0.954**
RMS y	**0.914**	**0.958**	**0.996**
total RMS	**1.309**	**1.313**	**1.379**

The number of examined objects is the same as in the quantitative analysis.

Note that the figures in Table 6-2 are based on image residuals. They show the difference of the automatically derived corner points and the digitized ones in the image. As our data sets were acquired with vertical viewing angles these results can be also interpreted as planimetric object space residuals.
But when dealing with images that were captured with oblique viewing angles, the buildings must be projected into object space in order to carry out a qualitative analysis in the reference system.

If we are dealing with stereo imagery, this transformation can be done by deriving the corners of a building in both of the images and then projecting them into object space. Epipolar constraints will support the derivation of the geometric building properties in the second image. The only information needed here is the interior and exterior orientation of both of the images.

In case the imagery is not available in stereo, but in mono together with an additional elevation source, we propose the following strategy:

- Project the building into object space by using a mean terrain height. At this point the building will not lie exactly on its true position in object space.
- Analyse the area around the projected building in object space and search for an elevation blob in the DSM. The search area is defined as

$h \cdot \tan(\alpha)$, where h is the average building height on the scene and α the sensor's viewing angle.

- Choose the maximum height of the DSM blob as reference elevation.
- Re-project the building corners from image space to object space, this time using the new reference elevation.

In our examples this "iterative projection" is not necessary, since the used imagery was acquired with nearly vertical viewing directions. But if this is not the case one has to be careful. If simply projecting the building corners onto an average elevation the following errors might occur.

Table 6-3: Projection errors with various constellations. (FoV of 1 degree approximately corresponds to high resolution satellite imagery (IKONOS, Quickbord), FoV of 60 degrees to airborne sensors (ADS40, HRSC).

Building Height in metres	Viewing Angle in degrees	Field of View in degrees	max. Planimetric Error in metres
5	0	1	0.04
10	0	1	0.09
25	0	1	0.22
5	10	1	0.93
10	10	1	1.85
25	10	1	4.63
5	20	1	1.87
10	20	1	3.74
25	20	1	9.35
5	0	60	2.89
10	0	60	5.77
25	0	60	14.43
5	10	60	4.20
10	10	60	8.39
25	10	60	20.98
5	20	60	5.96
10	20	60	11.92
25	20	60	29.79

Note, that Table 6-3 shows pure projection errors. If the building extraction errors should be included too, the errors of Table 6-2 multiplied with the according GSD must be added to the planimetric shifts of Table 6-3.

6.c Problems

This section discusses the weaknesses of the proposed strategies. A few examples are given for illustrating these drawbacks.

6.c.i Heterogeneous Roofs

Heterogeneity in the imagery can lead to misinterpretations and may mainly affect two steps of the proposed workflow:

- seed points determination
- adaptive region growing.

One of the basic assumptions made for finding seed points inside buildings was the homogeneity of roof colour and roof texture. In practice, there are cases with roofs of a rather heterogeneous, unsymmetrical behaviour. This is not a problem as long as the variations inside the building are less than the difference between the interior and exterior of the building.

Furthermore, also the behaviour of the adaptive region growing algorithm becomes unpredictable as it is unable to decide where to stop the growing process. In these cases additional constraints must be taken into consideration, like e.g. approximate region shape and size, in order to support the region growing algorithm.

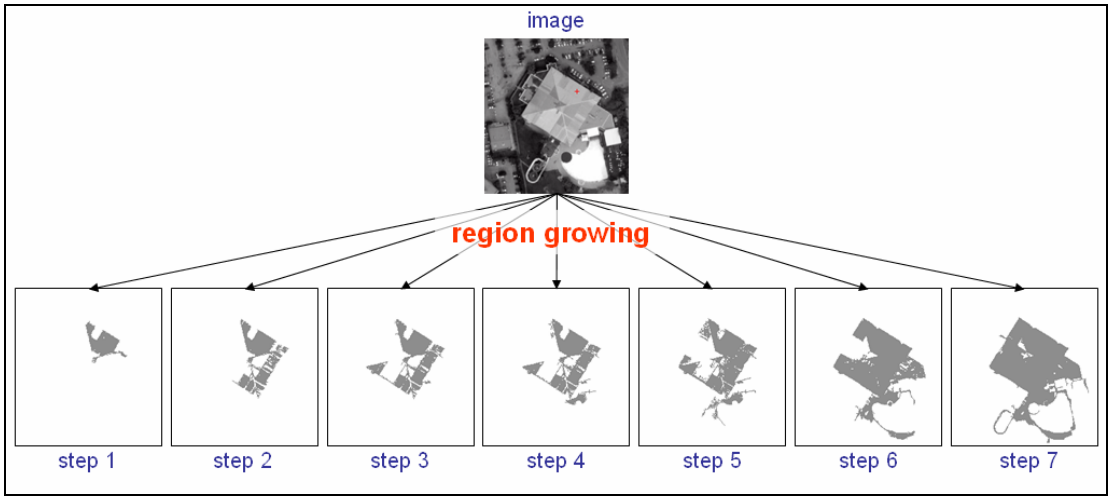

Figure 6-5: Example illustrating the region growing problem when dealing with heterogeneous roofs.

Such a problematic case is shown in Figure 6-5. When looking at the region growing steps, one can see that the accumulated area first flows over the border (from step 4 onwards) and not before step 7 the whole building is covered by the grown region. Even if the algorithm were able to automatically recognize when whole building is filled, the subsequent processes (i.e. edge and corner extraction) would fail to compute a correct solution.

6.c.ii Shadow Effects

Since the proposed approach is based on intensive image processing, effects leading to alterations in the grey value distribution can influence the result negatively.
Illumination problems or shadow effects can affect the image radiometry so that the exploitation techniques fail or produce erroneous results. Especially in highly urban areas with high-rise buildings and skyscrapers smaller buildings may be totally or partially covered with cast shadows (Figure 6-6). If the grey value difference over the building is very high, the algorithm cannot extract the building information correctly.

Figure 6-6: Highly urban area with big shadow effects (Seattle[3]).

6.c.iii Complex Buildings

4.d.v.B Level of Detail proved the capability to extract a rather high amount of detail of the buildings if and only if there is a clear distinction between the regions belonging to the building and not belonging to it.

In cases, were the building is very complex (Figure 6-7), the building outline becomes more and more difficult to be identified. In this example even human interpretation is difficult and it is hard to decide where the outer boundary of the building really is.

[3] Image taken from the LPS (ver.8.7) example data.

Figure 6-7: Complex building.

Figure 6-8 shows one example demonstrating the capability and limitations of the system. The grown region does not include the upper terrace and thus it is not included in the extracted vectors. The building is incomplete!

The problem is caused by the unclear arrangement of "outer" walls. Additional constraints and restrictions during the region growing process might help to overcome this insufficiency.

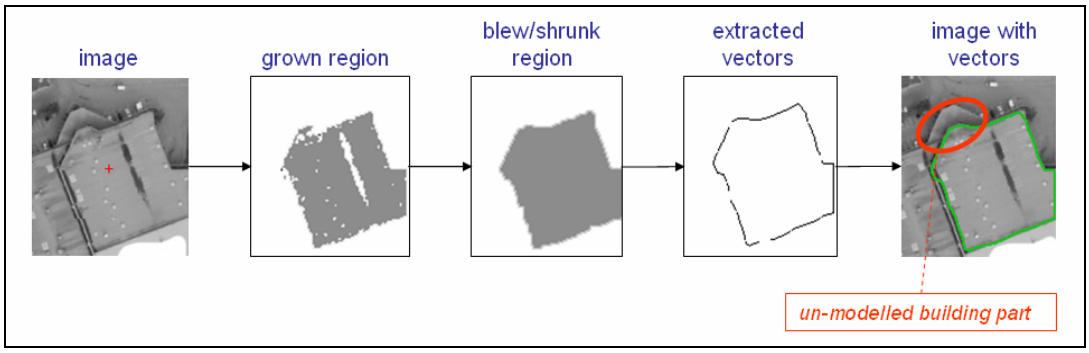

Figure 6-8: Example of incomplete building extraction.

6.c.iv DSM Inaccuracies

In *4.d Building Detection* it has been shown how nDSM information is handled in order to determine seed points. One can imagine that finding seed points might become difficult if the nDSM is not accurate or if it is extremely coarse compared to the building sizes.

Figure 6-9 is an extreme example to demonstrate the problem in case of a too coarse nDSM. It also made not much difference if the height threshold or kernel size (for homogeneity calculation) were chosen differently. A greater height threshold would cause the exclusion of buildings for subsequent steps, and a greater kernel size would cause regions over small buildings not to be considered as building

candidates and hence neglected further on. Though exaggerated, this example clearly shows the possible impact of too coarse DSMs on the quality building extraction.

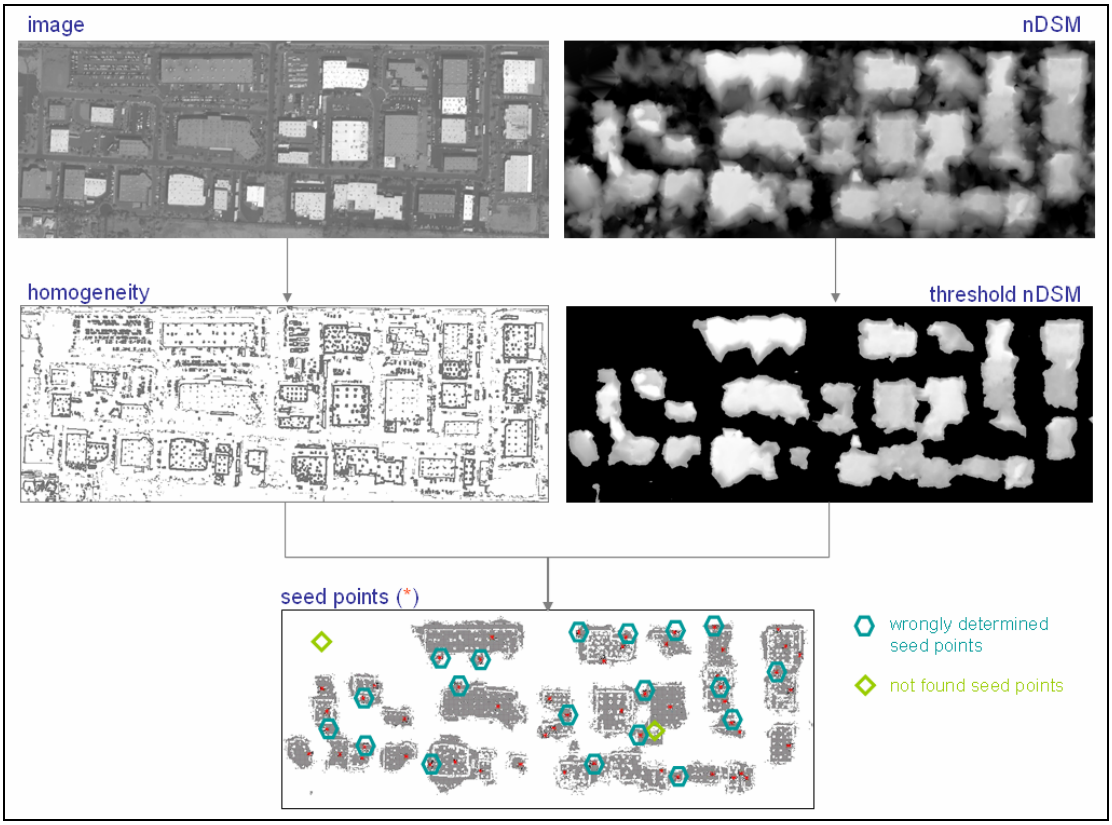

Figure 6-9: Wrong seed points due to nDSM inaccuracies (Quickbird subset of Denver).

6.c.v Adjacent Buildings

Adjacent buildings can produce two kinds of problems. The first one is shown in Figure 6-10 where the extracted buildings do not share a common border anymore. This misinterpretation is due to the fact that building outlines appear as prominent and rather thick ribbon-like features in the image. Hence the region growing algorithm stops too early and includes only the inner edge of these border ribbons. These errors can be removed by growing the found regions over their edges until the adjacent building has been reached. Of course, a-priori knowledge is necessary as one has to know in advance which building blocks are compound houses and where borders to adjacent buildings are located.

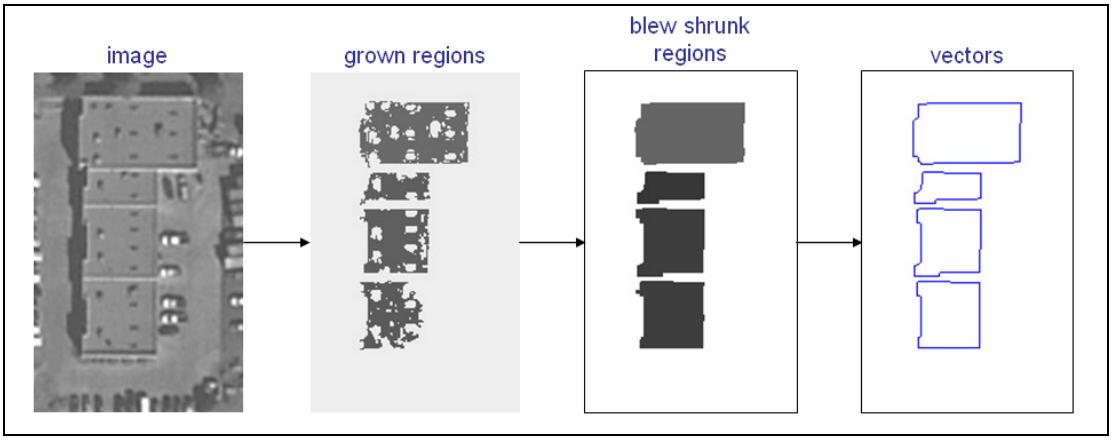

Figure 6-10: Buildings do not have a common border anymore.

The second, and more common drawback arises at adjacent buildings with tilted roofs. With one seed point on the building block the method would extract either just this part of the roof where the seed point has been set (Figure 6-11: step 1) or the whole block of buildings (Figure 6-11: step 5) from which this roof is just a small part. In either case the extraction would deliver wrong results.

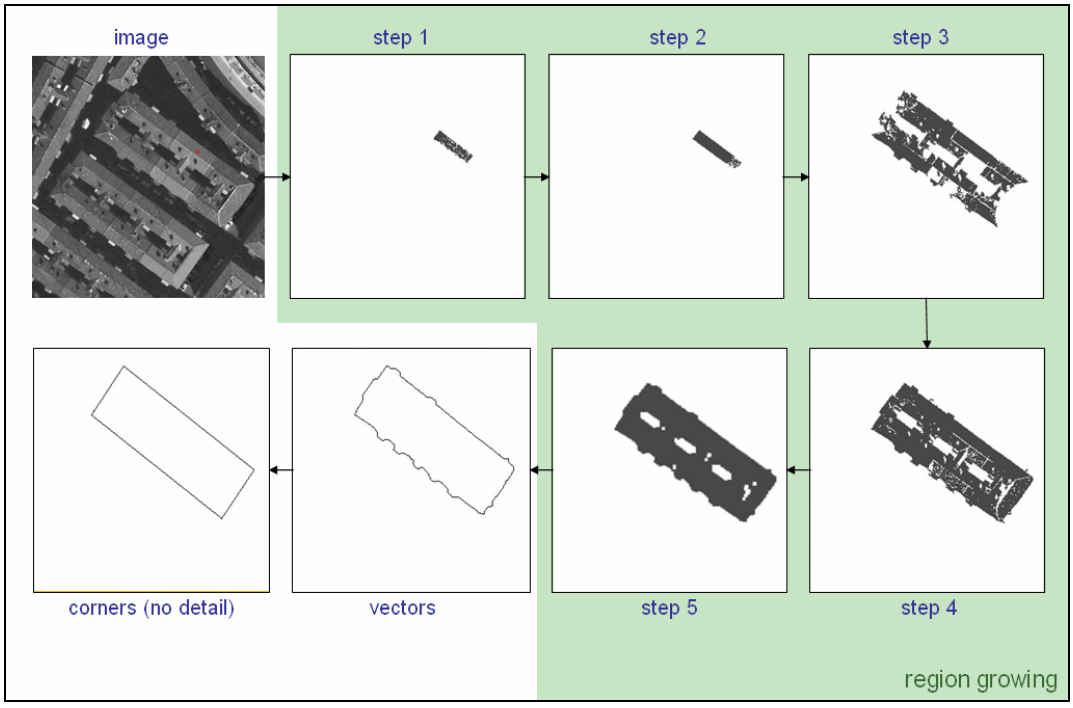

Figure 6-11: Problem of accumulation of neighbouring buildings (ADS40-Valladolid subset).

One suggested solution would be placing multiple seed points on a building block and then merging neighbouring roof parts that belong together. Again the major drawback is the fact that this information has to be given by the user and thus automation is impeded. This idea is illustrated for one house of a larger building block in Figure 6-12, where once three regions and then nine regions are merged in order to produce full and consistent vector data of a building.

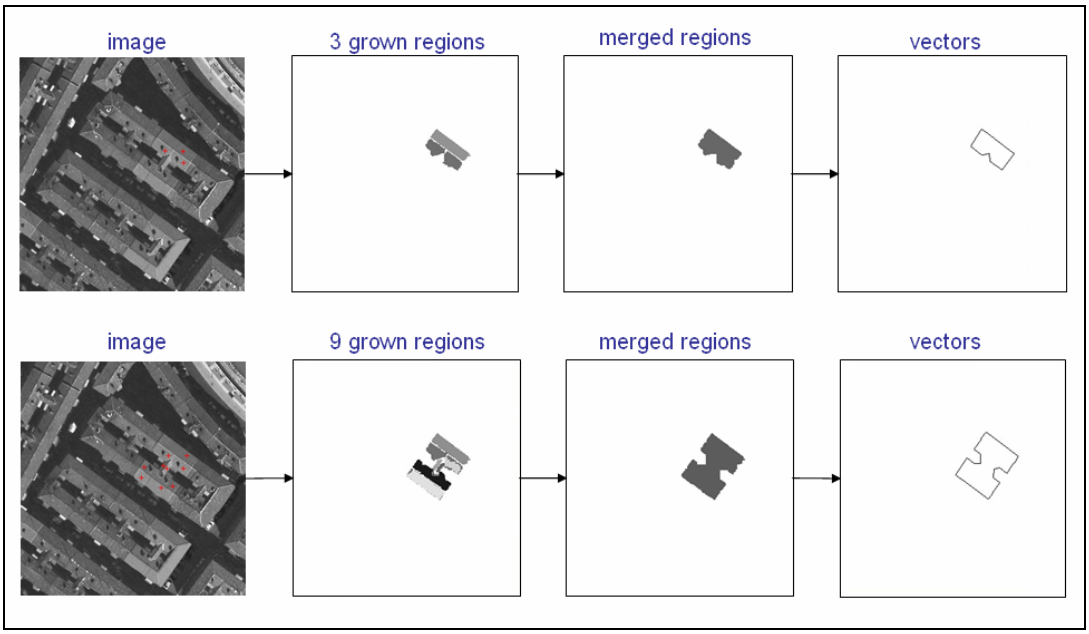

Figure 6-12: Merging adjacent roof parts (ADS40 subset of Valladolid).

7 Conclusions

The aim of this work was to propose a method for generating and updating Digital City Models which is based on images from spaceborne or airborne line scanning devices, on orthophotos if available and on elevation models.

While laser scanner data methods have become very common for DCM generation, imagery is not used often for this task. Many of these techniques are highly automated, but still far from mature. We tried to develop a method that provides higher flexibility and, although automated to a very degree, produces good results.

Various image processing techniques, such as Hough transformation, adaptive region growing, image matching, texture analysis, were employed and investigated for deriving the strengths and weaknesses of each. A variety of data sets were tested, coming from both spaceborne and airborne acquisition systems.

Many approaches employ colour images and hence can take advantage of multispectral analyses. Since we are dealing only with panchromatic imagery, a different technique was developed, making our research rather unique.

A general problem for DCM extraction algorithms is the detection of initial locations determining the positions of buildings on the scene. Most of the times this information is entered manually by the user or it is taken from e.g. digital cadastral maps.
To overcome this drawback an algorithm was developed for automatically finding the location of potential building candidates. It makes use of the panchromatic imagery and the corresponding DSM.

The implementation of the iterative Hough transformation provided the possibility to extract building edges very successfully. The method is very robust to noise and produced an interesting spin-off. The definition of a certain level of detail, just by

adjusting two parameters, allowed extracting buildings with full detail, but also on a very generalized level. This advantage can be used also for other application areas like cartographic mapping, where scale dependent generalization levels are required.

Figure 7-1 depicts a subset of a SPOT5 scene: the island of Salamis. Figure 7-2 illustrates various generalization levels that were automatically extracted by employing the proposed strategy. During the whole process no human interaction is necessary.

Another example is shown in Figure 7-3 and concerns the vectorization of classes in thematic maps. It is possible to reduce the number of corner points of an object without depriving the object's shape and detail. In Figure 7-3 an unsupervised classification (four classes) of a subset of a Landsat 7 scene (Neudiedlersee, Austria) was made and afterwards the vector data of the class lake was extracted. One can see that the accuracy and detail between the first three images is almost identical although the number of corner points is reduced by nearly one 10th.

Figure 7-1: Island Salamis (SPOT5-subset of Attica).

Figure 7-2: Multiple generalisation levels.

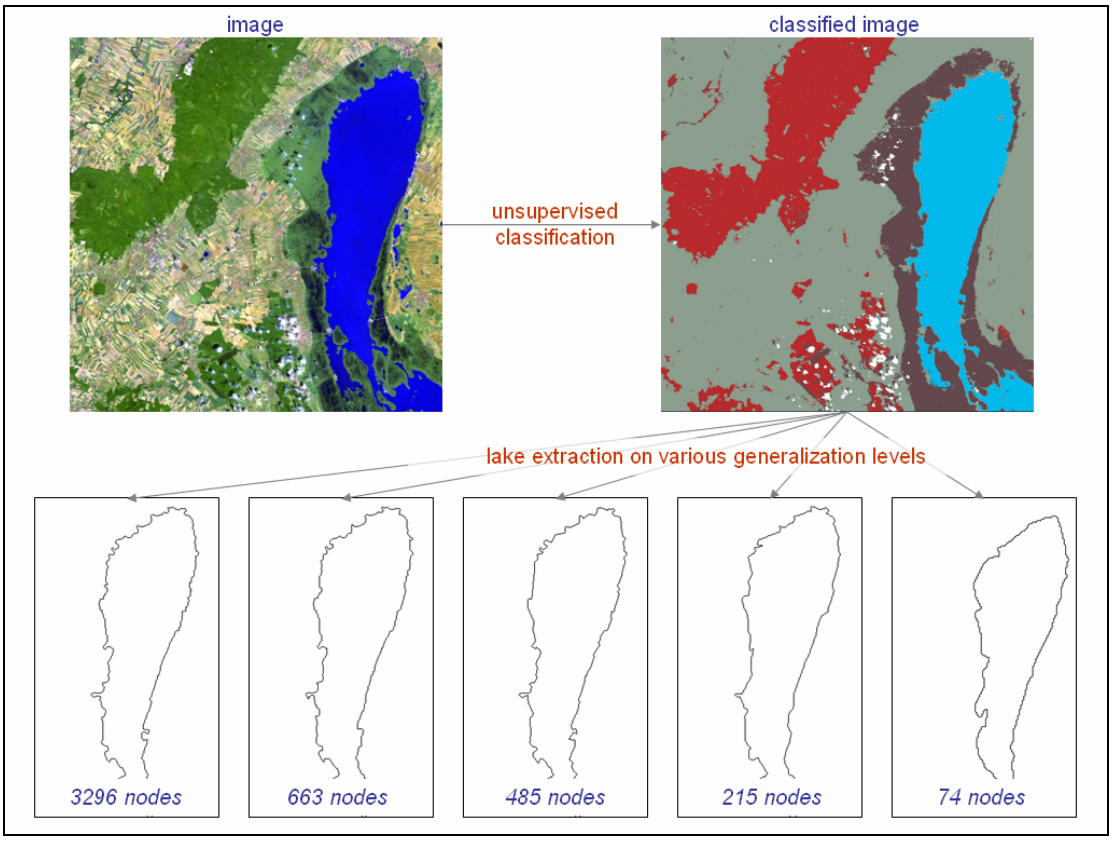

Figure 7-3: Lake extraction on multiple generalisation levels (Landsat7-subset of "Neusiedler See").

Through the research based on adaptive region growing and on the iterative Hough transformation we can conclude that the method is very powerful, but has also some weaknesses. One is the high dependence on the radiometric quality of the input imagery. Furthermore, rather small buildings will not be treated correctly. Thus we investigated two further methods (as add-ons) to see if we can overcome the previously mentioned problem. Image matching proved to be a very effective, but unfortunately a very time consuming method. The suggested strategy of texture analysis, although very efficient for pattern recognition over areas in small scale imagery, was not successful for extracting individual buildings.

The quality of a product is not only defined by its accuracy, but also by its reliability. Therefore quality analysis is necessary throughout the whole process and final product. These necessary steps were implemented and provide a good insight on quality during the whole building extraction process.

Through this research very good results were obtained, but nevertheless further investigations are necessary for improving the quality of the results even more.
Future work will be focused on:

- Extraction of objects with holes (e.g. houses with inner courtyards), i.e. deriving the inner and outer boundary of buildings.
- Research on constraint settings for aggregating neighbouring roof parts that belong to one building.
- Introduction of multispectral information for making the algorithms more efficient, especially as far as seed point determination is concerned.
- Extract edges on sub-pixel bases.
- Integrate a hierarchical approach in order to decrease computation time.

8 Acronyms

ADS40	Airborne Digital Sensor, three-line scanning system of company Leica Geosystems (www.gis.leica-geosystems.com/products/ads40/default.asp)
CCD	Charged Coupled Device, light sensitive semiconductor detector
CSG	Constructive Solid Geometry
DCM	Digital City Model
DLR	Deutsches Zentrum für Luft- und Raumfahrt, German Aerospace Center (www.dlr.de)
DoA	Degree of Artificiality, parameter for classifying RGB images into artificial (buildings, streets etc.) and natural (vegetation) objects
DSM	Digital Surface Model
DTM	Digital Terrain Model
GCP	Ground Control Point
GPS	Global Positioning System
GSD	Ground Sample Distance (see footnote on page 2)
HRSC	High Resolution Stereo Camera, digital camera based on the pushbroom scanner principle
IDL	Interactive Data Language, programming language developed by Research Systems Inc. (www.rsinc.com)
IFoV	Instantaneous Field of View, (see footnote on page 2)
IGD	International Geo Dynamics, surveying company in Athens, Greece (www.igd.gr)
IMU	Inertial Measurement Unit, registers orientation and position changes with extremely high accuracy

INS	Inertial Navigation System, see IMU
IPF	Institute of Photogrammetry and Remote Sensing of the Vienna University of Technology (www.ipf.tuwien.ac.at)
ISPRS	International Society of Photogrammetry and Remote Sensing (www.isprs.org)
ISTAR	Imagerie Stereo Appliquee au Relief, French mapping company, specialized in the processing of pushbroom scanner imagery (www.istar.fr)
JRC	Joint Research Centre (www.jrc.cec.eu.int)
LPS	Leica Photogrammetry Suite, software package by Leica Geosystems for photogrammetric processing of aerial and spaceborne imagery (www.gis.leica-geosystems.com/products/lps/)
MDL	Minimum Distance Length
MOLAND	Monitoring Land cover/use Dynamics (moland.jrc.it)
nDSM	normalized Digital Surface Model, difference of DSM minus DTM
NDVI	Normalized Difference Vegetation Index
PCA	Principal Component Analysis
RMS	Root Mean Square Error
RPCs	Rational Polynomial Coefficients, coefficients used for the orientation of high-resolution satellite imagery
SPOT Image	French company operating the SPOT family satellites (www.spotimage.fr)
TABU	Tool for Automated Building(s) Updating
TIN	Triangular Irregular Network

All URLs were last accessed on 20 May 2005.

9 References

Ackermann, F. (1984): Digital Image Correlation- Performance and Potential Application in Photogrammetry. Photogrammetric Record, Vol. 11, No. 64, pp. 429-439.

Adams, R. and Bischof, L. (1994): Seeded Region Growing. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 6, pp. 641-647.

Argialas, D. P. and Mavrantza, O. D. (2004): Comparison of Edge Detection and Hough Transform Techniques for the Extraction of Geologic Features. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B3, pp. 790-795.

Avrahami, Y., Raizman, Y. and Doytsher, Y. (2004): Semiautomatic 3D Mapping of Buildings from Medium Scale (1:40,000) Aerial Photographs. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B3, pp. 774-779.

Axelsson, P. (1999): Processing of Laser Scanner Data – Algorithms and Applications. ISPRS Journal of Photogrammetry and Remote Sensing. Vol. 54, Issues 2-3, pp. 138-147.

Baillard, C., Dissard, O., Jamet, O. and Maitre, H. (1998): Extraction and Textural Characterization of Above-Ground Areas from Aerial Stereo Pairs: A Quality Assessment. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 53, pp. 130-141.

Baillard, C. and Maitre, H. (1999): 3D Reconstruction of Urban Scenes from Aerial Stereo Imagery, A Focusing Strategy. Computer Vision and Image Understanding, Vol. 76, Issue 3, pp. 244-258.

Baillard, C., Schmid, C., Zissermann, A. and Fitzgibbon, A. (1999): Automatic Line Matching and 3D Reconstruction of Buildings from Multiple Views. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXII, Part 3, pp. 68-80.

Bartel, S., Bill, R. Boytscheff, C. and Koeniger, A. (1997): Datenfusion zur Erstellung Realistischer 3D-Geo-Informationssysteme fuer Staedtebauliche Planungen. Zeitschrift fuer Photogrammetrie und Fernerkundung (4), pp. 129-137.

Bauerhansl, C., Wuerlaender, R. and Rieger, W. (2004): Verbesserte Ableitung von Gelaendemodellen aus der Digitalen Bildzuordnung. Band 13 "Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation", Bochum , pp. 41- 49.

Bellmann, C. J. and Shortis, M.R. (2004): A Classification Approach to Finding Buildings in Large Scale Aerial Photographs. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B3, pp. 337-342.

Brenner, C. and von Goesseln, G. (2004): Tools and Workflow for the Rapid Acquisition of 3-D City Models. Proceedings of the 24th Urban Data Management Symposium, Information Systems and the Delivery of Social Benefits, Chioggia, Venice, on CD-rom.

Briese, C., Pfeifer, N. and Dorninger, P. (2002): Applications of the Robust Interpolation for DTM determination. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIV, Part 3A, pp. 55 – 61.

Boerner, A., Reulke, R., Scheele, M. and Terzibaschian, T. (1997): Stereo Processing of Image Data from an Airborne Three-Line CCD Scanner. Third International Airborne Remote Sensing Conference and Exhibition, Vol. I, pp. 423-431.

Cabral, P., Painho, M. and Gilg, J.P. (2004): Spatio-temporal Analysis of Urban Growth Patterns in Lisbon Metropolitan Area. Proceedings of the 24th Urban Data Management Symposium, Information Systems and the Delivery of Social Benefits, Chioggia, Venice, on CD-rom.

Carleer, A.P. and Wolff, E. (2004): Use of Spatial Information after Segmentation for Very High Spatial Resolution Satellite Data Classification of Urban Areas. Proceedings of the 24th Urban Data Management Symposium, Information Systems and the Delivery of Social Benefits, Chioggia, Venice, on CD-rom.

Chu, A., Sehgal, C.M. and Greenleaf, J.F. (1990): Use of Grey Value Distribution of Run Lengths for Texture Analysis. Pattern Recognition Letters, Vol. 11, Issue 6, pp. 415-419.

Christopher, J.S. and Warner, T.A. (2002): Scale and Texture in Digital Image Classification. Photogrammetric Engineering & Remote Sensing, Vol. 68, No. 1, pp. 51-63.

Collins, R., Jaynes, C., Cheng, Y., Wang, X., Stolle, F., Schultz, H., Riseman, E. and Hanson, A. (1998): The Ascender System: Automated Site Modelling from Multiple Aerial Images. Computer Vision and Image Understanding, Vol. 72, No. 2, November, 1998, pp. 143-162.

Dai, X.L. and Khorram, S. (1999): Remotely Sensed Change Detection based on Artificial Neuronal Networks. Photogrammetric Engineering & Remote Sensing, Vol. 65, No. 10, pp. 1187-1194.

Dial, G. and Grodecki, J. (2002): Block Adjustment with Rational Polynomial Camera Models. Proceedings of ACSM-ASPRS Annual Conference, on CD-rom.

Eastman, R. J., Mckendry, J.E. and Fulk, M.A. (1995): Change and Time Series Analysis. 2nd Edition in Explorations in Geographic Information Systems Technology, Volume I, United Nations Institute for Research and Training.

Ehlers, M., Schiewe, J. and Moeller, M. (2003): 3D City Modelling Using Ultra High Resolution and Multi Sensoral Remote Sensing. Zeitschrift fuer Photogrammetrie und Fernerkundung (6), pp. 30-37

Ehlers, M., Janowsky, R. and Gaehler, M. (2002): New Remote Sensing Concepts for Environmental Monitoring. Remote Sensing for Environmental Monitoring, GIS Applications, and Geology, Proceedings of SPIE Vol. 4545, 12 pages.

Eidenbenz, C., Kaeser, C. and Baltsavias, E. (2000): ATOMI – Automated Reconstruction of Topographic Objects from Aerial Images Using Vectorized Map Information. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIII, Part B2, pp. 462-471.

Eisenbeiss, H., Baltsavias, E., Pateraki, M. and Zhang, L. (2004): Potential of IKONOS and Quickbird Imagery for Accurate 3D Point Positioning, Orthoimage and DSM Generation. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B3, pp. 522-528.

Elmqvist, M., Jungert, E., Lantz, E., Persson, A. and Soderman, U. (2001):Terrain Modelling and Analysis Using Laser Scanner Data. International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 3/w4, pp. 219-227.

ERDAS (2003): ERDAS Field Guide™. 7th Edition, Leica Geosystems, Atlanta, Georgia, 698 pages.

Foerstner, W. (1982): On the Geometric Precision of Digital Correlation. International Archives of Photogrammetry and Remote Sensing, Vol. XXIV, No. 3, pp. 176 – 189.

Foerstner, W. (1995): Mid-level Processes for Automatic Building Extraction. Proceedings Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images, Monte Vertia, Switzerland, pp. 179-188.

Fraser, C. and Hanley, H. (2001): Geopositioning Accuracy of IKONOS Imagery: Indications from two Dimensional Transformations. Photogrammetric Record, Vol. 17, No. 98, pp. 317-329.

Fricker, P. (2001): ADS40-Progress in Digital Aerial Data Workflow. Photogrammetrische Woche 2001, pp. 105-116.

Fritsch, D. (1997): Experiences with the Airborne Three-Line Camera System DPA. Photogrammetrische Woche 1997, pp. 63-69.

Fuchs, C. (1998): Extraktion Polymorpher Bildstrukturen und ihre Topologische und Geomterische Gruppierung. Phd Thesis, Verlag der Bayrischen Akademie der Wissenschaften, Munich, 186 pages.

Fua, P. and Hanson, A.J. (1991): An Optimization Framework for Feature Extraction. Machine Vision Applications, Vol. 4, No. 2, pp. 59-87.

Gong, P., Marceau, D.J. and Howarth, P.J. (1992): A Comparison of Spatial Feature Extraction Algorithms for Land-Use Classification with SPOT HRV Data. Remote Sensing of Environment, Vol. 40, pp. 137-151.

Gonzalez, R.C. and Woods, R.E. (2002): Digital Image Processing, Second Edition. Patience Hall, New Jersey, 799 pages.

Granzow, E. (2001): Automatic Feature Recognition and Extraction from Remote Sensing Imagery. National Consortia on Remote Sensing in Transportation, Technical Notes, issue 5, 4 pages.

Gruen, A. (1985): Adaptive Least Square Correlation: A Powerful Image Matching Technique. South African Journal of Photogrammetry, Remote Sensing and Cartography, Vol. 14, No. 3, pp. 175-187.

Gruen, A. and Zhang, L. (2003): Automatic DTM Generation from Three-Line-Scanner (TLS) Images. Vol. XXXIV, Part 5/W10, on CD-rom.

Guelch, E. (2000): Digital Systems for Automated Cartographic Feature Extraction. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Part B2, pp. 241-256.

Haala, N. and Brenner, C. (1999): Virtual City Model from Laser Altimeter and 2D Map Data. Photogrammetric Engineering & Remote Sensing, Vol. 65, No. 7, pp. 787-795.

Haala, N. and Hahn, M. (1995): Data Fusion for the Detection and Extraction of Buildings. Proceedings Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhaeuser Verlag, pp. 211-220.

Hanley, H., Yamakawa, T. and Fraser, C. (2002): Sensor Orientation for High-Resolution Satellite Imagery. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIV, Part 1, on CD-rom.

Hannah, M.J. (1988): Digital Stereo Image Matching Techniques. American Society of Photogrammetry and Remote Sensing, Vol. 27, Part B3, pp. 280-293.

Haralick, R.M. (1979): Statistical and Structural Approaches to Texture. Proceedings IEEE, Vol. 67, No.5, pp. 786-804.

Haralick, R.M. and Shapiro, L. G. (1985): Image Segmentation Techniques. Computer Vision, Graphics, and Image Processing, Vol. 29, pp. 100-132.

Helava, U.V. (1988): Object-Space Least-Squares Correlation. Photogrammetric Engineering & Remote Sensing, Vol. 54, No. 6, pp. 711-714.

Henricsson, O. (1996): Analysis of Image Structures Using Colour Attributes and Similarity Relations. PhD Thesis, No. 11663, Swiss Federal Institute of Technology, 140 pages.

Hinz, A. and Doerstel, C. (2001): DMC - Systemkonzept und Ablauf der Datenverarbeitung. Photogrammetrie Fernerkundung Geoinformation, Vol. 3, pp. 190-198.

Hoffmann, C., Steinnocher, K., Kasanko, M., Toutin, T. and Cheng, P. (2001): Urban Mapping with High Resolution Satellite Imagery: IKONOS and IRS Data Put to the Test. GeoInformatics, Dossier, issue of December 2001, pp. 34-37.

Hofmann, O., Nave, P. and Ebner, H. (1982): DPS – A Digital Photogrammetric System for Producing Digital Elevation Models and Orthophotos by Means of Linear Array Scanner Imagery. International Archives of Photogrammetry and Remote Sensing, Vol. XXIV, Part 3, pp. 216-227.

Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P.J., Bunke, H., Goldgof, D.B., Bowyer, K., Eggert, D.W., Fitzgibbon, A. and Fisher, R.B. (1999): An Experimental Comparison of Range Image Segmentation Algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 7, pp. 673-689.

Hu, M.K. (1962): Visual Pattern Recognition by Moment Invariants. IRE Transactions on Information Theory, Vol. 8, pp. 179-187.

IDL (2004): Image Processing in IDL, Reference Manual. Research Systems Inc., Ver. 6.1, 560 pages.

Jain, A.K. (1989): Fundamentals of Digital Image Processing. First Edition, Englewood Cliffs, NJ: Prentice-Hall, 592 pages.

Jain, A.K. and Karu, K. (1996): Learning Texture Discrimination Masks. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 2, pp. 195-205.

Jaynes, C. (2000): Automatic Model Acquisition and Aerial Image Understanding. PhD Thesis, University of Massachusetts, 271 pages.

Jaynes, C., Riseman, E. and Hanson, A. (2003): Recognition and Reconstruction of Buildings from Multiple Aerial Images. Computer Vision and Image Understanding, Vol. 90, pp. 68-98.

Jensen, J. and Cowen, D.J. (1997): Principles of Change Detection Using Digital Remote Sensor Data. Integration of Geographic Information Systems and Remote Sensing. Cambridge Press, pp. 37-54.

Julesz, B. (1986): Texton Gradients: The Texton Theory Revisited. Biological Cybernetics, Vol. 54, pp. 247-251.

Jung, F. (2004): Detecting Building Changes from Multi-Temporal Aerial Stereopairs. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 58, Issues 3-4, pp. 187-201.

Kaufmann, R. and Seto, K. (2001): Change Detection, Accuracy, and Bias in a Sequential Analysis of Landsat Imagery in the Pearl River Delta. Agriculture, Ecosystems and Environment, Vol. 85(1-3), pp. 95-105.

Kemper, G., Calikoyan, M., Altan, O., Toz, G., Lavalle, C. and Demicelli, L. (2004): RS-techniques for Land Use Change Detection – Case Study of Istanbul. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B7, pp. 784-789.

Kilian, J., Haala, N. and Englich, M. (1996): Capture and Evaluation of Lasers Scanner Data. International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, Part B3, pp. 383-388.

Kim, C.E. and Anderson, T.A. (1984): Digital Disks and a Digital Compactness Measure. Proceedings of the 16th Annual ACM Symposium on Theory of Computing, pp. 117-124.

Kim, J.R. and Muller, J.-P. (2002): 3D Reconstruction from very high Resolution Satellite Stereo and its Application to Object Identification International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 4, on CD-rom.

Knudsen, T. and Olsen, B., (2003): Automated Change Detection for Updates of Digital Map Databases. Photogrammetric Engineering & Remote Sensing, Vol. 69, No. 11, pp. 1289-1296.

Kornus, W. and Lehner, M. (1999): Photogrammetric Point Determination and DEM Generation Using MOMS-2P7PRIRODA Three-Line Imagery, ISPRS Workshop on 'Sensor and Mapping from Space', on CD-rom.

Kraus, K. (1990): Fernerkundung, Band2, Auswertung Photographischer und Digitaler Bilder. Duemmler Verlag, Bonn, 614 pages.

Kraus, K. (1997): Photogrammetrie, Band1, Grundlagen und Standardverfahren. Duemmler Verlag, Bonn, 394 pages.

Kraus, K. (1996): Photogrammetrie, Band2, Verfeinerte Methoden und Anwendungen. Duemmler Verlag, Bonn, 488 pages.

Kraus, K. (2000): Photogrammetrie, Band3, Topographische Informationssysteme. Duemmler Verlag, Bonn, 419 pages.

Kreiling, W. (1976): Automatische Erstellung von Hoehenmodellen und Orthophotos durch Digitale Korrelation, PhD Thesis, Institute for Photogrammetry, University of Karlsruhe.

Kressler, F. and Steinnocher, K. (1996): Change Detection in Urban Areas Using Satellite Images and Spectral Mixture Analysis. International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences, Vol. XXXI, No. B7, pp. 379-383.

Leberl, F., Gruber, M., Ponticelli, M., Bernoegger, S. and Perko R. (2003): The Ultracam Large Format Aerial Digital System. American Society of Photogrammetry and Remote Sensing, Annual Conference, on CD-rom.

Lhomme, S., Weber, C., He, D-C., Morin, D. and Puissant (2004): Building Extraction from Very High Spatial Resolution Image. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B3, pp. 921-925.

Lohr, U. (2003): Digital Lagerichtige Orthophotos und LIDAR-Hoehenmodelle. Geo-Informations-Systeme GIS, Vol. 16, No. 3, pp. 26-29.

Lu, D. and Weng, Q. (2004): Spectral Mixture Analysis of the Urban Landscape in Indianapolis with Landsat ETM+ Imagery. Photogrammetric Engineering & Remote Sensing, Vol. 70, No. 9, pp. 1053-1062.

Maas, H.-G. (1999): The Potential of Height Texture Measures for the Segmentation of Airborne Laser Scanner Data. Proceedings of the Forth International Airborne Remote Sensing Conference, 21st Canadian Symposium on Remote Sensing, pp. 154-161.

Maas, H.-G. (2004): LiDAR: Very Interesting and Promising; Interview in GIM International, Vol. 18, No. 2, pp.6-9.

Matikainen, L., Hyyppä, J. and Kaartinen, H. (2004): Automatic Detection of Changes from Laser Scanner and Aerial Image Data for Updating Building Maps. International Archives of

Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B2, pp. 434-439.

Matikainen, L., Hyyppä, J. and Hyyppä, H. (2003): Automatic Detection of Buildings from Laser Scanner Data for Map Updating. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 3, pp. 218-224.

McDowell, J. (1997): The United Nations Registry of Space Objects. Harvard University, (*http://hea-www.harvard.edu/QEDT/jcm/space/un/un.html*, last accessed on 28 May 2003)

McGlone, J.C. and Schufelt, J. (1994): Projective and Object Space Geometry for Monocular Building Extraction. Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp. 54-61.

McIntosh, K. and Krupnik, A. (2002): Integration of Laser-Derived DSMs and Matched Image Edges for Generating an accurate Surface Model. ISPRS Journal of Photogrammetry and Remote Sensing, Volume 56, Issue 3, pp. 167-176.

Morain, S. and Baros, S.L. (1996): Raster Imagery in Geographic Information Systems. Onword Press, Santa Fe, 495 pages.

Mueller, K. and Segl, K. (1999): Object Recognition Based on High Resolution Panchromatic Satellite Imagery. Joint workshop of ISPRS on Sensors and Mapping from Space, on CD-rom.

Murakami, H., Nakagawa, H., Hasegawa, H., Shibata, T. and Iwanami, E. (1999): Change Detection of Buildings Using and Airborne Laser Scanner. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 54, Issues 2-3, pp. 148-152.

Nagao, M., Matsuyama, T. and Mori, H. (1979): Structural Analysis of Complex Aerial Photographs. Proceedings 6th International Joint Conference on Artificial Intelligence, pp. 610-659.

Nemmour H. and Chibani, Y. (2004): Combining Comparative and Simultaneous Analysis Approaches Based on Fuzzy Integration for Change Detection. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B3, pp. 709-713.

Neukum (2001): The airborne HRSC-AX cameras: Evaluation of the Technical Concept and Presentation of Application Results after one year Operations. Photogrammetrische Woche 2001, pp. 117-131.

Niederoest, M. (2000): Reliable Reconstruction of Buildings for Digital Map Revision. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIII, Part B3, pp. 635-642.

Olsen, B.P. (2004): Automatic Change Detection for Validation of Digital Map Databases. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B2, pp. 569-574.

Orun, A.B. (2004): Automated Identification of Man-Made Textural Features on Satellite Imagery by Bayesian Networks. Photogrammetric Engineering & Remote Sensing, Vol. 70, No. 2, pp. 211-216.

Pantazis, D., Yingcheng, L., Guo, T., Karathanassis, H., Stathakis, D. and Santimpantakis, C. (2004): The Update of Topographic Maps of Middle Scales Using High Resolution Satellite Images: Final Conclusions. Proceedings of the 24th Urban Data Management Symposium, Information Systems and the Delivery of Social Benefits, Chioggia, Venice, on CD-rom.

Pfeifer, N. (2003): Oberflaechenmodelle aus Laserdaten. Oesterreichische Zeitschrift fuer Vermessung und Geoinformation, Universitaetslehrgang – Laserscanning: Datenerfassung und Anwedungsorientierte Modellierung, Heft 4, pp. 243-252.

Puissant, A. and Weber, C. (2002): The Utility of Very High Spatial Resolution Images to Identify Urban Objects. Geocarto International, Vol. 17, No. 1, pp. 31-41.

Post, M.J. (1984): Minimum Spanning Ellipsoids. Annual ACM on Theory of Computing, pp. 108-116.

Roggero, M. (2001): Airborne Laser Scanning: Clustering in Raw Data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 3/w4, pp. 227-232.

Roggero, M. (2002): Object Segmentation with Region Growing and Principal Component Analysis. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 3, pp. 289-294.

Rottensteiner, F. (2001): Semi-Automatic Extraction of Buildings Based on Hybrid Adjustment Using 3D Surface Models and Management of Building Data in a TIS. Geowissenschaftliche Mitteilungen, Institut fuer Photogrammetrie und Fernerkundung, No. 56, 188 pages.

Rottensteiner, F. and Briese, C. (2002): A New Method for Building Extraction in Urban Areas from High-Resolution Lidar Data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 3A, pp. 295 - 301.

Rottensteiner, F., (2003). Automatic Generation of High-Quality Models from Lidar Data. Proceedings of IEEE, Computer Graphics and Applications, Vol. 23, No. 6, pp. 42-50.

Rottensteiner, F. and Schulze, M. (2003): Performance Evaluation of a System for Semi-Automatic Building Extraction Using Adaptable Primitives. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 3/W8, pp. 47-52.

Sandau, R. and Eckhardt, A. (1996): The Stereo Camera Family WAOSS/WAAC for Spaceborne/Airborne Appicatioins. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXI, Part B1, pp. 170-175.

Samadzadegan, F. and Abbaspour, A. (2004): Automatic Revision of Urban Geospatial Databases Using High Resolution Satellite Images Based Information Fusion Concepts. Proceedings of the 24th Urban Data Management Symposium, Information Systems and the Delivery of Social Benefits, Chioggia, Venice, on CD-rom.

Savopol, F. and Armenakis, C. (2004): Assessment of Lidar and Digital Camera Data in the Context of Rapid Change Detection Methodologies. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B4, pp. 968-973.

Schickler, W. (1992): Feature Matching for Outer Orientation of Single Images Using 3D Wireframe Controlpoints. International Archives of Photogrammetry and Remote Sensing, Vol. XXIX, Part B3, pp. 591-598.

Schickler, W. and Poth, Z. (1996): The Automatic Interior Orientation and its Daily Use. International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, Part B3, pp. 746-751.

Scholten, F. and Gwinner, K. (2003): Band 12 "Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation", Bochum, pp. 419-426.

Sindhuber, A. (1998): Ergaenzung und Fortfuerhung eines Digitalen Landschaftmodelles mit Multispektralen und Hochaufloesenden Fernerkundungsdaten. Geowissenschaftliche Mitteilungen, Heft 48, Institut fuer Photogrammetrie und Fernerkundung, 82 pages.

Singh, A. (1989): Digital Change Detection Techniques using Remotely-Sensed Data. International Journal of Remote Sensing, Vol. 10, No. 6, pp. 989-1003.

Sirinyildiz, T. (2004): Change Detection Analysis by Using IKONOS Imagery. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B7, pp. 1268-1270.

Sithole, G. (2001): Filtering of Laser Altimetry Data Using a Slope Adaptive Filter. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 3/w4, pp. 203-210.

Sohn, G. and Dowman, .I.J. (2002): Terrain Surface Reconstruction by the Use of Tetrahedron Model with the MDL Criterion. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 3A, pp. 336-344.

SPOT (2004): Access: The Key to Increased Data Use. Spot Magazine - Focus, No. 37, pp. 22-23.

SPOT (2003): Map Updating: As a Valuable Aid to Economic Development. Spot Magazine - Special Report, No. 36, pp. 12-14.

Steinle, E., Oliveira, F.H., Baehr, H.-P. and Loch, C. (1999): Assessment of Laser Scanning Technology for Change Detection in Buildings. Proceedings XVIIth CIPA Symposium, WG4, 7 pages.

Taillandier, F. and Deriche, R. (2004): Automatic Building Reconstruction from Aerial Images: A Generic Bayesian Framework. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B3, pp. 343-348.

Tavakoli, M. and Rosenfeld, A. (1982): Building and Road Extraction from Aerial Photographs. IEEE Transactions on Systems, Man and Cybernetics, Vol. 12, No. 1, pp. 84-91.

Trinder, J.C., Wang, P., Lu, Y., Wang, Z. and Cheng, E.D. (2002): Aspects of Digital Elevation Model Determination. . International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. XXXIV, Part 4, on CD-rom.

Ulm, K. (2003): Improved City Modelling with Cybercity-Modeler (CC-Modeler^{ZM}) Using Aerial-, Satellite Imagery and Laserscanner Data. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 5/W10, on CD-rom.

Vapnik, V. (1995): The Nature of Statistical Learning Theory. New York, Springer Verlag, 188 pages.

Venkateswar, V. and Chellapa, R. (1990): A Framework for Interpretation of Aerial Images. Proceedings 10th International Conference on Pattern Recognition, Vol. 1, pp. 204-206.

Voegtle, T. and Steinle, E. (2003): On the Quality of Object Classification and Automated Building Modelling Based on Laserscanning Data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 3/W13, pp. 149-155.

Vosselman, G. (2000): Slope Based Filtering of Laser Altimetry Data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIII, Part B3, pp. 935-942.

Vosselman, G., Gorte, B.G.H. and Sithole, G. (2004): Change Detection for Updating Medium Scale Maps Using Laser Altimetry. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B3, pp. 207-212.

Vozikis, G., Jansa, J. and Fraser, C. (2003): Alternative Sensor Orientation Models for High-Resolution Satellite Imagery. Band 12 "Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation", Bochum , pp. 179- 186.

Vozikis, G. (2004): Urban Data Collection: An Automated Approach in Remote Sensing. Proceedings of the 24th Urban Data Management Symposium, Information Systems and the Delivery of Social Benefits, Chioggia, Venice, on CD-rom.

Vozikis, G. (2004): Automated Generation and Updating of Digital City Models Using High-Resolution Line Scanning Systems. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B7, pp. 1033-1038.

Wang, Y. (1998): Principles and applications of structural image matching. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 53, Issue 3, pp. 154-165.

Weeks, A. (1996): Fundamentals of Electronic Image Processing. Wiley-IEEE Press, 576 pages.

Weidner, U. and Foerstner, W. (1995): Towards Automatic Building Extraction from High-Resolution Digital Elevation Models. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 50, Issue 4, pp. 38-49.

Westin, T. (1990): Precision Rectification of SPOT Imagery. Photogrammetric Engineering & Remote Sensing, Vol. 56, No. 2, pp. 247-253.

Wewel, F. and Scholten, F. (1999): High Resolution Stereo Camera (HRSC): Multispectral Data Acquisition and Photogrammetric Data Processing. Proceedings of the 4th International Airborne Remote Sensing Conference and Exhibition, Ottawa, pp. 263-271.

Wu, C. and Doerschuk, P. (1994): Texture Based Segmentation Using Markov Random Fields. Natural and Stochastic Methods in Image and Signal Processing III, Vol. 2304, No. 16, pp. 86-93.

Zhang, C., Baltsavias, E. and Gruen, A. (2001): Knowledge-Based Image Analysis for 3D Road Reconstruction. Asian Journal of Geoinformatics, Vol. 1, No. 4, pp. 3-14.

Zhang, Y. (2001): A Spectral and Spatial Information Integrated Approach for Tree Extraction from High-Resolution Digital Imagery. Digital Earth, Fredericton, Canada, 9 pages.

Ziou, D. and Tabbone, S. (1997): Edge Detection Techniques – An Overview. Department de Mathematiques et Informatique, Université de Sherbrooke, Québec, Canada, pp. 1-41.