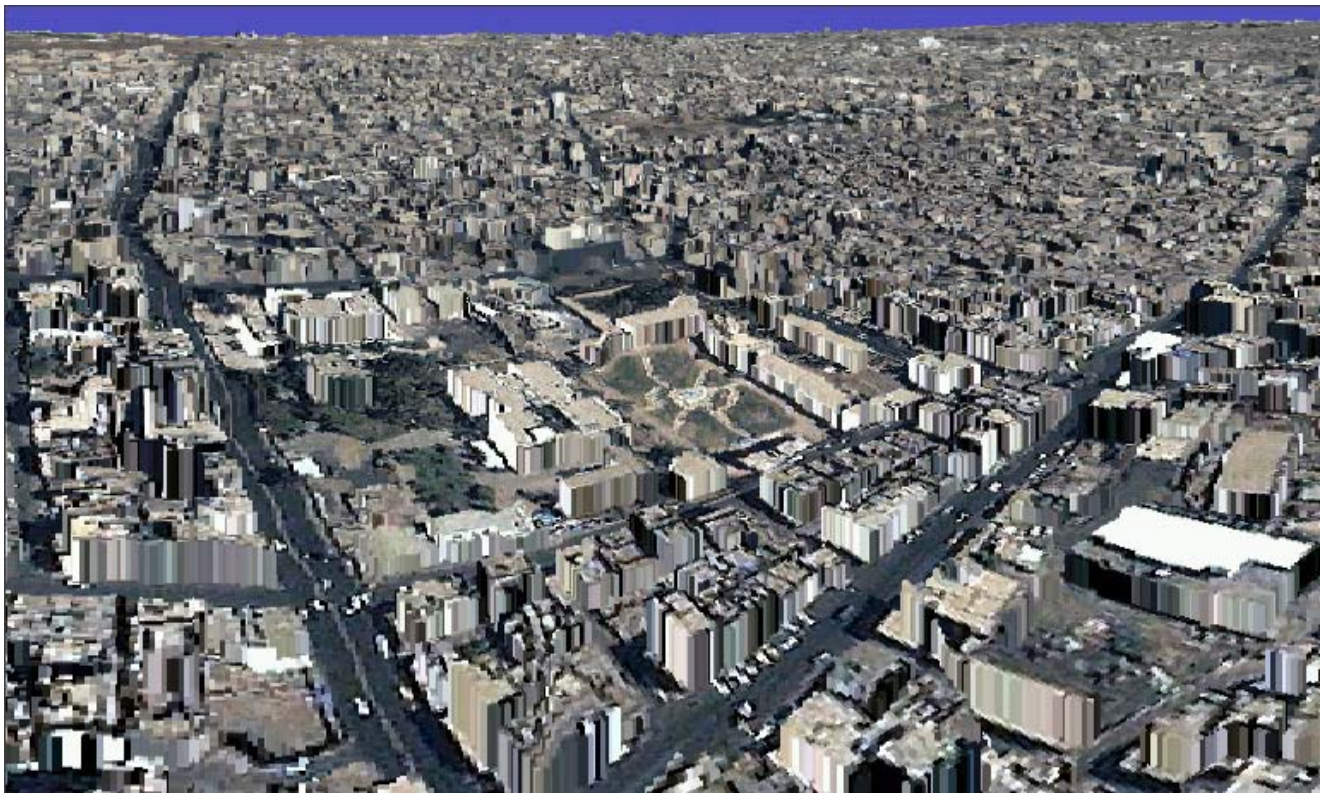
JRC Scientific and Technical Reports

3D-Building Height Extraction from Stereo IKONOS Data

Quantitative and Qualitative Validation of Digital Surface Models
Derivation of Building Height and Building Outlines

Sandra Eckert



EUR 23255 EN - 2008

The Institute for the Protection and Security of the Citizen provides researchbased, systems-oriented support to EU policies so as to protect the citizen against economic and technological risk. The Institute maintains and develops its expertise and networks in information, communication, space and engineering technologies in support of its mission. The strong crossfertilisation between its nuclear and non-nuclear activities strengthens the expertise it can bring to the benefit of customers in both domains.

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Address: European Commission – Joint Research Centre
Institute for the Protection and Security of the Citizen – Support to External Security
TP 267, Via E. Fermi, 2749
21020 Ispra, (VA), Italy
E-mail: sandra.eckert@jrc.it
Tel.: +39 0332 789651
Fax: +30 0332 785154

http://ses.jrc.it
http://ipsc.jrc.ec.europa.eu
http://www.jrc.ec.europa.eu

JRC 43067

EUR 23255 EN
ISBN 978-92-79-05127-2
ISSN 1018-5593

DOI: 10.2788/68063

Luxembourg: Office for Official Publications of the European Communities

Printed in Luxembourg

Cover: 3D visualization of Sana'a. IKONOS rgb composite draped over the final DSM.

Abstract

This report is dealing with the digital surface model generation from VHR stereo satellite data with the focus on building height and shape extraction. The report provides a theoretical insight into orthorectification methods based on either empirical or rigorous, physical models and the theoretical aspects of digital surface model extraction. Orthorectification of stereo satellite data highly influences the accuracy of a digital surface model besides the selected matching methodology applied during the surface model extraction process. The requirement and ideal distribution of ground control points is discussed. In the final part of the report the results of four software packages, ENVI, PCI Geomatica, RSG and Leica Photogrammetric Suite, tested for urban DSM generation, are presented and described.

The orthorectification accuracy analyses were done using QuickBird and IKONOS data. For the digital surface model accuracy analyses stereo IKONOS data were mainly used. The data is commercially purchasable and is the satellite data with the highest geometric resolution that can easily be acquired as stereo datasets. Two datasets were used to perform the tests. One study area is situated in Nairobi where a variety of building types are present, from high-rise buildings to small illegal shacks. The second study area is in Graz, which was mainly chosen because a very detailed reference surface model was available.

The orthorectification accuracy results for both test areas show that the rigorous physical model performed best with an accuracy below pixel size with RMSE of 0.31m (x-direction) and 0.45m (y-direction) for the Nairobi QuickBird dataset. The empirical rational function based orthorectification achieved RMSE larger than the pixel size of 0.60m. A 1-order polynomial adjustment resulted in slightly better accuracies than a 0-order polynomial adjustment.

The orthorectification for the Graz IKONOS dataset resulted in similar accuracies: with the rigorous model RMSEs of 0.56m (x-direction) and 1.06m (y-direction) were achieved. The rational function based orthorectification with 1-order polynomial adjustment resulted in RMSEs of 0.97m and 0.63m. However, large RMSE of more than 10m for one of the used ground control points indicates that the model is not stable for the entire test area.

The physical models proved to be stable for both IKONOS and QuickBird data using few GPCs, eight for Nairobi, or a large number of GCPs, 26 for Graz. Consequently it is recommended to use the rigorous physical model for orthorectification of VHR satellite data if GCPs are available and a high geometric accuracy of the data is necessary. The RF model is still a viable alternative when high accuracy GCPs are very limited or not available. If the topography is rather flat in a dataset a 0-order polynomial adjustment with RF model orthorectification might be sufficient but a 1-order polynomial adjustment should be preferred if the terrain is rugged.

In the Nairobi test area mainly qualitative analyses and pointwise quantitative analyses were performed due to lack of reference data e.g. building heights and/or building outlines or a high-resolution digital surface model. The generated DSMs were evaluated by comparing them with reference height data taken from the internet, the GPS ground elevation data collected in the field, and data calculated by an alternative height extraction methodology. Additionally, two qualitative tests were conducted to come to a conclusion in terms of DSM quality relating to building height and shape extraction.

The five evaluation tests have shown that all tested software packages created DSMs that performed well in at least one of the tests. They all have advantages and disadvantages. Height accuracy as well as clear building shape extraction is of great importance for the use of DSMs in information extraction for settlement analysis and mapping. The highlighted tests are representing these criteria best. Judging them it can be concluded that overall the PCI and RSG software performed best. They should be favoured for DSM extraction. However, a big disadvantage of RSG is the computation time. Still, both software packages, PCI and RSG are recommended for urban DSM extraction.

The quantitative accuracy assessment for the test area of Graz has shown that the best vertical estimation results were achieved with the software packages of LPS and PCI followed by RSG. The vertical MAE for built-up and impervious areas was 2.20m for PCI, 2.28m for LPS and 2.55m for RSG respectively. The RMSE was 3.05m, 2.96m and 3.25m respectively. Besides the vertical also the horizontal error should be considered depending on the different orthorectification methodologies applied (rigorous physical or RF based models). The shift compared with the reference DSM was between 3.06m and 3.27m. However, the qualitative, visual DSM evaluation has not confirmed the quantitative results. LPS with the best quantitative accuracy created fuzzy building outlines and contains low details in areas with smaller objects. PCI and RSG both produced DSMs with clear building outlines. They both are able to extract high details in areas with small buildings. Besides achieving the largest error in the quantitative analysis due to an erroneous mountain in the North of the test area ENVI also had problems in extracting correct multi-storey buildings outlines in the denser city area. It achieved good visual results with high details for rather small buildings.

Summarizing it can be said that the recommendations made for the Nairobi test area were confirmed by the quantitative and qualitative accuracy assessment done in the Graz test area. Both, PCI and RSG performed well or achieved at least acceptable results in both the quantitative and qualitative analysis. They both are recommended for digital surface model generation over built-up areas and settlements. Although LPS achieved the best quantitative accuracy it failed in creating DSMs with high details and clear building outlines. This is essential for building height extraction if the building outlines have to be still extracted from the data itself and are not available from cadastral offices as vector data layer. At last, ENVI showed a weak performance in creating large erroneous elevations in the test areas. It failed in achieving good quantitative results for both buildings and ground elevations. However, it should be mentioned that it successfully extracted fine structures of very small buildings in the Nairobi test area.

Two problems have to be addressed to extract building heights from stereo satellite data. First, the object height information has to be derived from the generated DSM. Two methodologies were presented to derive the object height layer: an indirect and a direct methodology. Second, the building outlines have to be delineated and extracted. A possible approach was proposed based on watershed segmentation.

The first results of the two tested methodologies are promising. A mean absolute error of 4.53m and 5.97m respectively was achieved when comparing them with reference building heights. Medium-height buildings were estimated well with an approximate error of one floor. Tall buildings are estimated with larger errors of two or more floors. These discrepancies have to be further analyzed. In case they are constant or linear the addition of an offset could be integrated into the current extraction methodologies.

Additionally, a building outline extraction approach based on watershed segmentation and preliminary results were presented. The methodology successfully detected most buildings. However, problems occur where buildings have complicated outlines. The extracted shapes of most building outlines are approximated and not representing the generally rectangular shapes of buildings. The next working steps will focus on the improvement of these approaches.

Table of Contents

List of Figures

List of Tables

Acronyms

CBD	Central Business District
DEM	Digital Elevation Model
DSM	Digital Surface Model
DTM	Digital Terrain Model
DGPS	Differential Global Positioning System
GCP	Ground Control Point
GIS	Geographical Information System
GMOSS	Global Monitoring for Security and Stability
ICP	Independent Check Point
IFOV	Instantaneous Field of View
LPS	Leica Photogrammetric Suite
MAE	Mean Absolute Error
PAN	Panchromatic
RF	Rational Function
RMSE	Root Mean Square Error
RPC	Rational Polynomial Coefficient
RSG	Remote Sensing Software Package Graz
SAR	Synthetic Aperture Radar
SPOT-HRV	SPOT High Resolution Visible
SRTM	Shuttle Radar Topographic Mission
VHR	Very High Resolution

1 Rationale

The work described herein focuses on the development of information layers on built-up areas in support of territorial management risk and damage assessment in urban areas of mega cities in Africa. The tested and/or developed methodologies are based on photogrammetric height extraction from optical stereo satellite data. The goal is to extract very detailed digital surface models and the building height information available within the surface model, to delineate single buildings, to calculate the number of floors and, at a later stage, define the type of building and estimate the number of population.

This work is conducted within the Information Support for Effective Rapid External Action (ISFEREA) project of the Joint Research Centre. The work is in support of the European Commission External Relations services that include DG External Relations, DG Development, DG AIDCO, DG Enlargement and ECHO in general. This work is also relevant for UN emergency agencies that include UNHCR, WFP, WHO active in post disasters when information on built-up structures become very relevant, UN Habitat and other international organization such as the World Bank.

The technical work supports policies related to crisis management, humanitarian aid and development programs. The research contributes to address vulnerability and risk analysis, long term city planning and development as well as rapid emergency response and support to reconstruction.

2 Introduction

Recently the need for 3D data describing built-up areas has increased. The availability of commercial very high resolution (VHR) optical satellite sensors such as IKONOS and QuickBird offer the possibility of stereo satellite data acquisition. Besides obtaining very detailed and up-to-date imagery of an area they allow the extraction of the third dimension and thus the generation of digital surface models (DSMs). DSMs can be used to extract information about the present built-up structures and affected population and be of great use in disaster management, damage, vulnerability and risk assessment analysis or urban planning.

Diverse research has been done dealing with height information extraction from stereo VHR satellite data since their availability. The research focused on different specific tasks such as on the development of physical models for improved orthorectification, or on DSM generation and the improvement of their accuracy (Eisenbeiss, et al., 2004, Toutin, 2001) with the focus on the ground control point distribution. Only few studies exist that directly compare a large variety of software packages and their DSM extraction modules. Al-Rousan and Petrie, 1998, validated five software packages and their performance generating DSMs and orthoimages from SPOT stereo pairs. Kay and Zielinski recently evaluated the accuracy of DSMs generated from Cartosat stereo data using two different software packages and Poon et al. compared three different software packages generating DSMs from IKONOS stereo pairs.

Another field of research that has received increasing attention with the increased availability of optical VHR data and laser scanning data is the conversion of DSMs into digital terrain models (DTMs). A variety of methodologies has been developed based on a) interpolation (Kraus and Pfeifer, 1997, Champion and Boldo, 2006), b) advanced filtering (Vosselman, 2000), or c) mathematical morphology (Binard, et al,, 2006, Dell'Acqua et al., 2001, Zhang, et al., 2003, Arefi and Hahn, 2005) in order to filter or remove objects such as trees, cars, and buildings from the dataset. In our research we attempt to develop building height extraction methods by modifying and improving the previous methodologies and testing them.

The presented research in this report focuses on:

1. the comparison of orthorectification methodologies for VHR satellite data

2. the comparison of the performance of a selection of remote sensing software packages and their DSM generation modules and a qualitative and quantitative accuracy assessment of the generated DSMs

3. the presentation and evaluation of two building height extraction methodologies

4. the development of a building outline extraction approach based on watershed segmentation.

The research was done in two test areas, Nairobi and Graz. For both test areas a set of stereo IKONOS data and GPS points were available. For the quantitative evaluation a highly accurate reference DSM was available for the Graz test area. For the Nairobi test area only pointwise quantitative evaluation measurements were available. Additionally, a qualitative evaluation was done for both test areas.

3 The Importance of 3D information

Height information is mainly integrated in five application fields within ISFEREA that are dealing with settlements: **risk assessments** to natural disasters and **post disaster damage assessments**, **building height and population estimations**, and **urban planning** and **support to reconstruction**.

Aerial photography, fine scale maps and increasingly VHR satellite imagery are used to quantify risk to natural disaster – and if a natural hazard strikes and disaster unfolds the same data are used to estimate the severity of the damage based on the nature and the intensity of the disaster. These data sources are also used by civil protection and homologous institutions in developing countries. Unfortunately the valuable information is under the control of ministry of interior and only a number of countries, especially in high income countries have started to make the data available on a commercial basis. The humanitarian community attempting to address mass emergencies in low income countries may often have had difficulties in accessing these geo-spatial data when available. The availability of VHR imagery virtually anywhere in the world is contributing towards enhancing response to disasters. VHR imagery has the advantage to **give an overview of the affected area shortly after a disaster**, especially in case of inaccessibility.

At present, global elevation information can be derived only from the globally available SRTM data at 90 m resolution. However the spatial resolution of SRTM data often doesn't fulfil the requirements of local hazard phenomena. An alternative is fine scaled topographic maps which have to be digitalized but they are often not available in developing countries or remote areas. Therefore, surface elevation models automatically derived from stereo VHR data can **provide highly detailed topography information** capturing even small variations in topography that may be critical for natural hazards, such as landslides, lahars, lava flows or flash floods. Satellite imagery is **increasingly used in risk assessments** addressing three parameters of the risk equation. (1) The risk to **natural hazards** for those hazards where **topography** is an aggravating factor; (2) the "stock of built up areas" or the entire **physical infrastructure** and (3) the **physical vulnerability of buildings**. Besides extracting the surface height information which includes object or building heights, VHR data can be used to assess the "stock of built up areas" by extracting the number of buildings and multiplying it with the building height. If the physical characteristics of the buildings additionally are extracted from the satellite data it can be used to calculate physical vulnerability to natural hazards. However, the information on physical vulnerability remains highly scene dependent. By and large every built up area has its own specificity that are determined by climate, topography, culture, affluence and technology in the very period the buildings have been constructed. The analysis of physical vulnerability will therefore remain very case specific. However, building structures especially the most technologically advanced have common construction standards. 3D information on buildings can greatly improve the understanding on their vulnerability. In fact, most of small, informally built buildings tend to be the most fragile.

The knowledge of the **number of floors of a building** and the **utility of the buildings** can be used to **estimate the population**. This can be helpful in case of a disaster, where not only the physical damage can be analyzed, by **comparing pre-disaster imagery with post-disaster imagery**, but also the number of **affected population**.

After a disaster, single VHR data can be a support to urban planning and reconstruction monitoring. In reconstruction monitoring, post-disaster imagery is often compared with very recent imagery acquired a few months or a year after the start of the reconstruction process.

These examples show that 3D information is increasingly influencing and improving geospatial analysis. It is either made available by labour-intensive map digitization and surveying, or by means of satellite based photogrammetry which can provide a viable and cost-effective alternative to get height information on settlements, especially in case of area inaccessibility, unavailability from other sources or time constraints.

4 Theory

4.1 Orthorectification

4.1.1 Theoretical Background

Each image acquisition system (Figure 1) produces unique geometric distortions in its raw images and consequently the geometry of these images does not correspond to the terrain or to a specific map projection of end-users. Obviously, the geometric distortions vary considerably with different factors such as the platform (airborne versus satellite), the sensor (optical or SAR; low to very high resolution), and also the total field of view. However, it is possible to make general categorizations of these distortions.

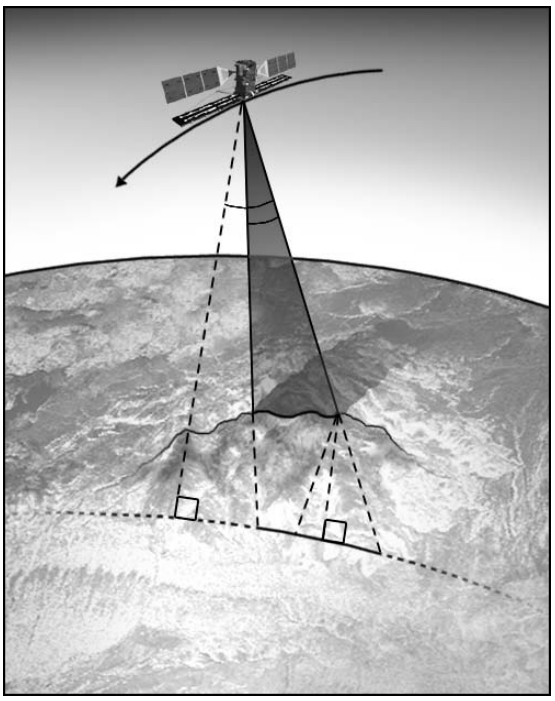


Figure 1: Geometry of viewing of a satellite scanner in orbit around the Earth.

The sources of distortion can be grouped into two broad categories: the *Observer* or the acquisition system (platform, imaging sensor and other measuring instruments, such as gyroscope, stellar sensors, etc.) and the *Observed* (atmosphere and Earth). In addition to these distortions, the deformations related to the map projection have to be taken into account because the terrain and most of GIS end-user applications are generally represented and performed respectively in a topographic space and not in the geoid or a referenced ellipsoid. Table 1 describes in more detail the sources of distortion for each category and sub-category.

Category	Sub-Category	Error Sources
The Observer	Platform	Variation of the movement Variation in platform attitude Variation in sensor mechanics
Acquisition System	Sensor	Viewing/look angles
	Measuring instruments	Panoramic effect with field of view Time-variations or drift Clock synchronicity
The Observed	Atmosphere	Refraction and turbulence
	Earth	Curvature, rotation topographic effect
	Map	Geoid to ellipsoid Ellipsoid to map

Table 1: Description of error sources for the two categories, the Observer and the Observed with the different sub-categories (Toutin, 2004)

The geometric distortions of Table 1 are predictable or systematic and generally well understood. Some of these distortions, especially those related to the instrumentation, are generally corrected at ground receiving stations or by image vendors. Others, for example those related to the atmosphere, are not taken into account and corrected because they are specific to each acquisition time and location and information on the atmosphere is rarely available.

The remaining distortions associated with the platform are mainly orbit and Earth related (quasi-elliptic movement, Earth gravity, shape and movement) (Escobal 1965, Centre National d'Études Spatiales 1980, Light et al. 1980). Depending of the acquisition time and the size of the image, the orbital perturbations have a range of distortions. Some effects include:

- platform altitude variation in combination with sensor focal length, Earth's flatness and terrestrial relief can change the pixel spacing;
- platform attitude variation (roll, pitch and yaw) can change the orientation and the shape of VIR images; it does not affect SAR image geometry; and
- platform velocity variations can change the line spacing or create line gaps/overlaps.

The remaining sensor-related distortions include:

- calibration parameter uncertainty such as in the focal length and the instantaneous field of view (IFOV) for VIR sensors or the range gate delay (timing) for SAR sensors; and
- panoramic distortion in combination with the oblique-viewing system, Earth curvature and topographic relief changes the ground pixel sampling along the column.

The remaining Earth-related distortions include:

- Rotation, which generates latitude-dependent displacements between image lines;
- Curvature, which for large width image creates variation in the pixel spacing; and
- Topographic relief, which generates a parallax in the scanner direction.

The remaining deformations associated with the map projection are:

- the approximation of the geoid by a reference ellipsoid; and
- the projection of the reference ellipsoid on a tangent plane.

All these remaining geometric distortions require models and mathematical functions to perform geometric corrections of imagery: either through 2D/3D empirical models (such as 2D/3D polynomial or 3D rational functions, RFs) or with rigorous 2D/3D physical and deterministic models. With 2D/3D physical models, which reflect the physical reality of the viewing geometry (platform, sensor, Earth and sometimes map projection), geometric correction can be performed step-by-step with a mathematical function for each distortion/deformation or simultaneously with a "combined" mathematical function. The step-by-step solution is generally applied at the ground receiving station when the image distributors sell added-value products (georeferenced, map oriented or geocoded) while the end users generally use and prefer the "combined" solution (Toutin, 2004).

4.1.2 Empirical Models - Rational Functions

The 2D/3D empirical models (e.g. rational functions) can be used when the parameters of the acquisition systems or a rigorous 3D physical model are not available. Since they do not reflect the source of distortions described previously, these models do not require *a priori* information on any component of the total system (platform, sensor, Earth and map projection).

The interest in 3D rational functions (RFs) has been renewed in the civilian photogrammetric and remote sensing communities due to the launch of the first civilian very high-resolution sensor, IKONOS, in 1999. Since sensor and orbit parameters were not included in the meta-data. Image vendors thus provide with the image all the parameters of 3D RFs. Consequently, the users can directly process the images without GCP for generating orthoimages with DEM, and even post-process to improve the RF parameters with GCPs. This approach was adopted by two resellers. They provide RF parameters for IKONOS Geo images (Grodecki 2001) and QuickBird-2 images (Hargreaves and Roberston 2001) using 3rd-order RF parameters. Since biases or errors still exist after applying the RFs, the results need to be post-processed with few precise GCPs (Fraser et al. 2002) or the original RF parameters can be refined with linear equations requesting more precise GCPs (Lee et

al. 2002).

Recent studies using 3D RFs with different very high resolution images (level-1A EROS-A1, IKONOS Geo, level-1A QuickBird-2) showed inferior and less consistent results (Toutin et al. 2002, Kristóf et al. 2002) than orthorectifications based on physical models. Some inconsistencies and errors with the IKONOS orthoimages generated from RFs were not explained (Davis and Wang 2001) while these errors did not appear using the 3D physical model. Tao and Hu (2002) achieved 2.2-m horizontal accuracy with almost 7-m bias while processing stereo IKONOS images using 1st-approach RF method. Kristóf et al. (2002) and Kim and Muller (2002) obtained 5-m random errors computed on precise independent check points (ICPs) when using respectively, precise GCPs to compute the RFs or the RFs provided with the stereo-images and a post-processing with GCPs to remove the bias. Larger errors away from the GCPs were also reported (Petrie, 2002). Since the academic results are not entirely confirmed by the 'end-users' results, more research should be thus performed to evaluate the true applicability and the limitations of these 3D RFs for high-resolution images in an operational environment and also in any study site, especially with high relief (Toutin, 2004).

4.1.3 Physical Model

2D/3D physical functions are used to perform the geometric correction difference, depending on the sensor, the platform and its image acquisition geometry. Although each sensor has its own unique characteristics, one can drawn generalities for the development of 2D/3D physical models, in order to fully correct all distortions described previously. The physical model should mathematically model all distortions of the platform (position, velocity, attitude for optical sensors), the sensor (viewing angles, panoramic effect), the Earth (ellipsoid and relief for 3D) and the cartographic projection. The geometric correction process can address each distortion one by one and step by step or simultaneously (Toutin, 2004).

4.1.4 Data Processing Level

The raw 'level 1A' images are preferably to be used with 3D physical models because they are derived from co-linearity equations which are well known and developed. Since different 3D physical models are largely available for such VIR images, raw 1A-type images should be favoured by the remote sensing community. However, IKONOS and QuickBird-2 data are generally systematically corrected and georeferenced and thus the 'level 1B' images just retain the terrain elevation distortion, in addition to a rotation-translation related to the map reference system. The map-oriented images ('level 2A') also retain the elevation distortion but image lines and columns are no more related to sensor-viewing and satellite directions. However, a 3D physical model has been still approximated and developed for IKONOS Geo images using basic information of the metadata and celestial mechanics laws (Toutin and Cheng 2000). Even approximated, this 3D physical model ('using a global geometry and adjustment') has been proven to be robust and to achieve consistent results over different study sites and environments (urban, semi-rural, rural, Europe, North and South America), different relief (flat to high) and different cartographic data (DGPS, ortho-photos, digital topographic maps, DEM) (Toutin 2003a).

4.1.5 GCP Requirements Depending on the Choice of Model

Whatever geometric model used, even the empirical RF approach will require some GCPs to remove the bias or refine RF parameters to compute/refine the parameters of the mathematical functions in order to obtain a cartographic standard accuracy. Generally, an iterative least-square adjustment process is applied when more GCPs than the minimum number required by the model (as a function of unknown parameters) are used. The number of GCPs is a function of different conditions: the method of collection, the sensor type and resolution, the image spacing, the geometric model, the study site, the physical environment, GCP definition and accuracy and the final expected accuracy. The weakest aspect in GCP collection, which is of course different for each study site and image, will thus be the major source of error in the error propagation and overall error budget of the bundle adjustment.

Since empirical models do not reflect the geometry of viewing and do not filter errors, many more GCPs than the theoretical minimum are required to reduce the propagation of input errors in the geometric models. However, when using RF models with the already-computed parameters provided by the image vendor, few (1-10) GCPs are only needed to remove the bias or to refine the RF parameters. When more than one image is processed, each image requires its own GCPs and the geometric models are generally computed separately (no relative orientation or link between adjacent images). However, some block adjustment can be performed with RFs (Dial and Grodecki 2002, Fraser et al. 2002). Since empirical models are sensitive to GCP distribution and number, GCPs should be spread over the full image(s) in planimetry and also in the elevation range for the

3D models to avoid large errors between GCPs. It is also better to have medium-accurate GCPs (lakes, tracks, ridges) than no GCP at the tops of mountains. If the image is larger than the study site it is recommended to reduce the GCP collection to the study site area because the empirical models only correct locally.

With 3D physical models, fewer GCPs (1-6) are required per image. When more than one image is processed a spatio-triangulation method with 3D block-bundle adjustment can be used to process all images together (optical and SAR). It enables users to drastically reduce the number of GCPs for the block with the use of tie points (TPs) (Belgued et al. 2000, Kornus et al. 2000, Toutin 2003b, c, d). When the map and positioning accuracy is of the same order of magnitude as the image resolution, twice (or a little less) the theoretical minimum is recommended. When the accuracy is worse, the number should be increased depending also of the final expected accuracy (Savopol et al. 1994). Since more confidence, consistency and robustness can be expected with physical models (global image processing, filtering input errors) than with empirical models, it is not necessary to increase the number of GCPs in operational environments. GCPs should preferably be spread at the border of the image(s) to avoid extrapolation in planimetry, and it is also preferable to cover the full elevation range of the terrain (lowest and highest elevations). Contrary to empirical models, it is not necessary to have a regular distribution in the planimetric and elevation ranges. Since the physical models correct globally the GCP collection has to be performed in the full image size, even if the study site is smaller. First, it will be easier to find GCPs over the full image than over a sub-area and more homogeneity is thus obtained in the different area of the image. GCP cartographic co-ordinates to correct IKONOS or QuickBird-2 data should have an accuracy of at least 0.5 m which can be obtained with differential GPS measurements.

4.2 Digital Surface Model Extraction from Stereo VHR Data

4.2.1 Theoretical Background

Using satellite data to produce DSMs and DEMs has been a rather recent development research topic. Efforts, based on digital matching and additional image processing techniques, have been made in digital photogrammetry to develop automatic methods for surface reconstruction (Baltsavias and Stallmann, 1993, Baillard and Dissard, 2000). Software packages have been developed by several companies to reconstruct the surface and automatically generate DSMs using optical stereo satellite data.

The different processing steps using stereo images can be described in broad terms as follows:

1) to acquire the stereo image data with supplementary information such as ephemeris and attitude data if available;
2) to collect GCPs to compute or refine the stereo model geometry;
3) to extract the elevation parallax by automatic matching techniques;
4) to compute the 3-D cartographic coordinates using 3-D stereo intersection; and
5) to create and possibly postprocess the DSM (smoothing, filtering, 3-D editing, etc.)

4.2.2 Ground Control Point Collection

Geometric modelling solutions, which have been adapted to suit the geometry of scanner imagery, employ the well-known co-linearity and co-planarity equations. With parametric modelling, few GCPs are required. In an operational environment their number will vary as a function of their accuracy and they should preferably be spread at the border of the stereo pair to avoid extrapolation in planimetry and cover the full elevation range of the terrain. Different types of GCPs can be used:

- full control points with known XYZ co-ordinates;
- tie points with unknown cartographic co-ordinates.

The last type is useful to reinforce the stereo geometry and fill in gaps where there is no XYZ-GCP. Furthermore, GCPs displayed only on one image in or outside the stereo pair can also be acquired as complementary points to the stereo GCPs. Combined with tie points they can also help to avoid extrapolation in planimetry in areas where there is no stereo GCP (Toutin, 2001). The final accuracy of the stereo geometry is mainly dependent on the GCPs cartographic and image coordinates.

4.2.3 Extracting Elevation Parallax

In digital photogrammetry elevation parallax is extracted through image matching. When working with airphotos, computer-assisted visual matching on analytical stereo-workstations is principally used, and where digital images are the basis, automated image matching is common. Before elevation extraction, the images are generally resampled into epipolar or quasi-epipolar geometry to remove the y-parallax. Any matching procedure for determining homologous points can be restricted to the rows of the images which results in smaller search space. Another methodology to reduce the number of false matches and multiple solutions is to apply hierarchical techniques where different levels of resolution of the image are being created and the match is applied in the highest level of the pyramid in order to provide an initial approximation. The two main matching techniques are area based and feature based matching.

The idea in area based matching is to compare the grey level distribution of a small image patch, with its counterpart in the other image. The template is the image patch which usually remains in a fixed position in one of the images. The search window refers to the search space within which the image patches are compared with the template (Schenk, 1996). Image matching can be performed with several methods of which cross-correlation is considered to be the most accurate (Leberl, et al., 1994) and is largely used with remote sensing images (Toutin, 2001).

Cross-correlation techniques have a long tradition for finding conjugate points in photogrammetry. The idea is to measure the similarity of the template with the matching window by computing the correlation factor. The cross-correlation factor is determined for every position *r,s* of the matching window within the search window. Next is to determine the position *u, v,* which yields the maximum correlation factor. If the search window is constrained to the epipolar line, then the correlation factors can be plotted in a graph, the maximum is found by fitting a polynomial through the correlation values (Schenk, 1996).

Feature based matching comprises two stages. First, interesting features and their attributes in all images are detected and second the corresponding features are determined. The features are extracted in each image individually prior to matching them. Local features are points, edges and lines, and regions. Larger features are called structures. Larger (global) features are called structures. Global features are usually composed of different local features. Besides the attributes of the local features, relations between these local features are introduced to characterize global features. These relations can be geometric such as the angle between two adjacent polygon sides or the minimum distance between two edges, radiometric such as the difference in grey value or grey value variance between two adjacent regions or topologic, such as the notion that one feature is contained in another. Matching with global features is also referred to as relational or structural matching (Shapiro and Haralick, 1987). Features should be distinct with respect to their neighbourhood, invariant with respect to geometric and radiometric influences, stable with respect to noise, and seldom with respect to other features. Different operators have been developed to extract these features. Afterwards a cross-correlation or least-squares matching is applied.

In the tested software packages only area based matching techniques are applied.

4.3 Satellite Image Software Packages

The DSM comparison analyses were performed with four different software packages offering a DSM generation module. Three of them are commercially available, *PCI Geomatica (PCI), ENVI and Leica Photogrammetric Suite (LPS)*. The fourth software package is a non-commercial product developed at Joanneum Research, Institute of Digital Image Processing, Graz; it is called *Remote Sensing Software Graz (RSG)*.

4.3.1 ENVI 4.3

In ENVI the relationship between image space and ground space for IKONOS stereo data is modelled through rational polynomials. The software requires the RPCs provided with the stereo images. Additionally, GCPs can be collected. The stereo images are then reprojected into epipolar projection in order to remove the y-parallax and thus reduce the search space for finding corresponding image points during the automatic image matching. For DSM extraction several parameters can be set such as the terrain detail (a maximum of seven scans), the terrain relief type (low, moderate, high), the matching window size (5x5 up to 15x15 pixels) and the minimum correlation value to be accepted. During the matching process the previously defined “number of scans” are performed. For each scan the original images are being resampled to multiples of the original pixel size. The

extracted DSM at the resampled pixel size is then used as a start value for the subsequent cross-correlation scan. ENVI provides possibilities for interpolating and filtering the raw DSM.

4.3.2 PCI Geomatica 9.1.7

PCI Geomatica offers not only the rational polynomial function to establish a stereo-model but also a rigorous, physical model. The rational polynomials provided by the data provider can be optimized by adding GCPs.

The rigorous, physical model was originally developed to suit the geometry of push-broom scanners, such as SPOT-HRV and subsequently adapted as an integrated and unified geometric modelling to geometrically process multi-sensor images (Toutin, 1995). This model applied to different image types is robust and not sensitive to GCP distribution as soon as there is no extrapolation in planimetry and elevation. The geometric modelling represents the well-known collinearity condition (and coplanarity condition for stereo-model), and takes into account the different distortions relative to the global geometry of viewing. The model has been adapted to IKONOS images by taking into account the image characteristics and the available information in the metadata file. More details on the mathematic model and development (collinearity equations), its applicability and full results with a large data set of IKONOS images can be found in Toutin, 2003a.

Once the GCPs are collocated in both images for either the rigorous or the RF-based approach the geometric model is computed. The rigorous approach requires a minimum of six GCPs. Then the images are reprojected into an epipolar projection. The *x* and *y* displacement between each matching pixel in the two input images is determined by taking a small image chip centred around a particular pixel in one image and moving it around on the second image until the best local match is found. The size of the image chip and matching area is predefined by the software and can not be changed. The procedure uses a hierarchical sub-pixel normalized cross-correlation matching method to find the corresponding pixels in both images. The difference in location between the images gives the disparity or parallax arising from the surface relief, which is converted to absolute elevation values above the local mean sea level datum using a 3D space intersection solution (Toutin and Cheng, 2001). The DSM can be calculated at different sampling intervals. Additionally, a DSM detail parameter (high, medium, low) defines how much detail will be included in the extracted DSM. Before the completion of the final DSM the software provides an interpolation algorithm and several filters to post-process the raw DSM.

4.3.3 RSG 5.1.2

With the RSG software the stereo IKONOS scenes are map projected using the rational polynomials provided by the data provider. The rational polynomials can be optimized using linear add-on polynomials being determined from GCPs. RMS, minimum and maximum values of point residuals, resulting from a backward transformation of control points into the image are used to conclude on the accuracy of the rational polynomial optimization. For matching of the stereo scenes RSG uses an extended version of the “feature vector” matching method. It was developed at Joanneum Research, Institute of Digital Image Processing, Graz (G. Paar and W. Pölzleitner, 1992, M. Caballo-Perucha, 2003). The components of the “feature vectors” are in general represented by various convolution and variance filters or other suchlike features (H. Raggam et al., 2005). Several parameters can be adjusted for the matching step such as the search window size, the number of pyramids as well as the maximum back-matching distance. One essential feature is the cross correlation coefficient, which formerly used to be applied for image matching as such. Back-matching is used in order to get a reliability feed back for the individual matching results. Based on the results of the image matching, the digital raster surface model is then generated. It is achieved via 3D point intersection of the projected rays defined by the matched image points. In this step unreliable matching results, which will lead to wrong ground coordinates can be rejected. Subsequently, a regular raster surface model is interpolated. The gaps which may still be inherent to this raster model due to unreliable point rejection are interpolated using a versatile interpolation mechanism (H. Raggam et al., 2005). In Table 6 the applied parameter settings for the tested software packages are listed.

4.3.4 LPS 8.7

In Leica Photogrammetry Suite (LPS) the relation between ground and image space can be defined either by the use of a RF model, by a number of polynomial equations or a physical model. For the adjustment of the physical model GCPs are required. LPS supports most of the metadata describing the physical model of the system provided with satellite imagery but also the respective rational function model (e.g., QuickBird,

IKONOS). It is possible to refine the RF model data based on existing GCP and additionally tie points can be collected automatically, identifying corresponding points in both scenes of the stereo pair. In order to reduce the search area the images are projected in epipolar geometry before the matching process.

LPS makes use of correlation and hierarchical image matching to automatically extract DSMs by using an interest operator. A series of interest points or feature points is identified on each image in a block and then matched. The cross-correlation coefficients are calculated for each correlation window among the search window. A correlation window exists on the reference image and a search window exists on the neighbouring overlapping image. An interest point located on the reference image may have more than one possible match on the adjacent overlapping images. For each set of possible image points identified by LPS a correlation coefficient is computed.

Strategy parameters (e.g. search and matching/correlation window size, correlation coefficient limit, etc.) can be defined depending on the stereo image characteristics, additionally there are eight predefined scanning strategies which all differ in the size of the search window, size of the correlation window, the correlation coefficient limit but also the amount of DSM filtering, the topographic type and the object type. Search window size, correlation window size, and correlation coefficient limit may be adjusted automatically if the corresponding checkbox is enabled in the Set Strategy Parameters dialog. If adaptive change is selected, LPS computes and analyzes the terrain features after each pyramid and sets the strategy parameters accordingly.

Once the correlation coefficient has been computed for each set of possible matching image points, various statistical tests are used within LPS to determine the final set of image points associated with a ground point on the surface of the Earth. Once the final set of image points has been recorded, the 3D coordinates associated with the ground feature are computed. The resulting computation creates a DSM mass point. A mass point is a discrete point located within the overlap portion of at least one image pair, and whose 3D ground coordinates are known. A technique known as space forward intersection is used to compute the 3D coordinates associated with a mass point.

5 Test Areas & Datasets

5.1 Nairobi

5.1.1 Study Area

Nairobi is located at 1° 16' S and 36° 48' E. The altitude in the study area varies between 1520m and 1800m above sea level (asl). The Ngong hills, located to the west of the city are the most prominent topographical feature of the Nairobi area besides Nairobi River and its tributaries.

Nairobi is the most populous city in East Africa with an estimated urban population in the wider Nairobi Area of between three and four million. According to the 1999 census the administrative area of Nairobi 2'143'254 inhabitants lived within 684 km^2 (Central Bureau of Statistics. http://www.cbs.go.ke, last visited: 17.12.2007). It is estimated that Nairobi's population will reach 5 million in 2015 (Ethos international, http://www.ethosinternational.org/site/pp.asp?c=dkLQK4MNItG&b=493007, last visited 17.12.2007).

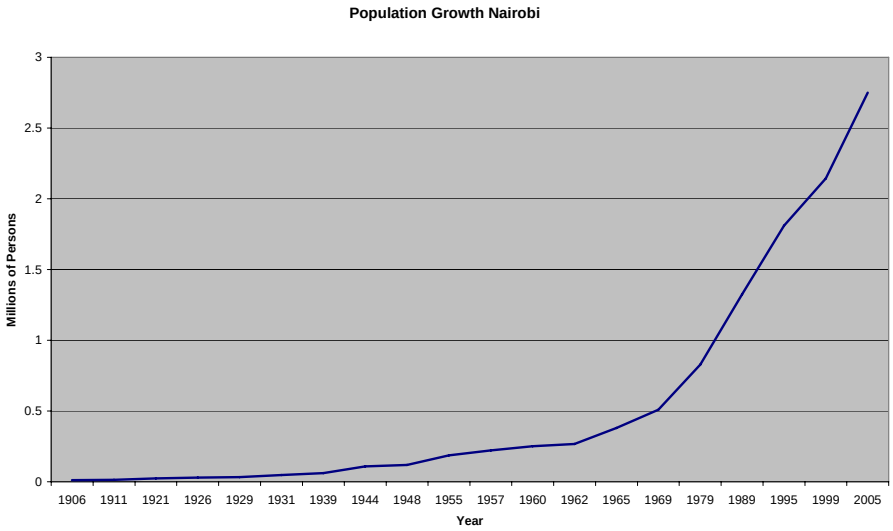


Figure 2: Population growth of Nairobi (Olima, 2001)

Half of the population have been estimated to live in slums which cover just 5% of the city area. The growth of these slums is a result of urbanization, poor town planning and the unavailability of loans for low income earners. One of the largest slums in Africa, Kibera, is situated in the southwest of Nairobi. The slum covers two square kilometres and is on government land. Other notable slums include Mathare and Korogocho. Altogether, 66 illegal settlements are counted as slums within Nairobi. Identifying them and analyzing their vulnerability with the help of very high resolution satellite imagery is a big challenge but could be of great support to urban planning authorities and international donors.

5.1.2 Data

For the Nairobi test area two VHR remote sensing datasets were acquired. Additionally, 17 differentially measured high-precision GPS points were collected. They are well distributed over the test area and have an accuracy of 0.5m or better. They were used for the orthorectification of the QuickBird data which from then on served as georectified reference for the orthorectification and DSM generation of the stereo IKONOS dataset.

Sensor	Level	Resolution vis/nir [m]	Acquisition Date	Sensor Elev. [°]
QuickBird	Standard 2A	0.6/2.4	14.01.2003	
Ikonos (left)	Standard stereo epipolar	1.0/4.0	15.12.2003	82.80754
Ikonos (right)	Standard stereo epipolar	1.0/4.0	15.12.2003	66.91180

Table 2: Overview of remote sensing data available over Nairobi.

Graz

5.1.3 Study Area

The second test area is in the city of Graz in Austria where a highly accurate reference DSM dataset is available. The tested subset comprises an area of 2km by 1km and represents buildings typically found in European cities such as multi-storey buildings with centre courtyards, large industrial buildings, residential row houses and single residential houses. However, there are no small buildings, such as shacks, present in Graz like in the Nairobi study area. The elevation in the Graz study area ranges from 390m asl to 480m asl., rising from West to East.

5.1.4 Data

The stereo IKONOS dataset used for the quantitative accuracy assessment was acquired in June 2007. The stereo satellite scene was orthorectified with 23 differential GPS points. Table 3 gives an overview of the scene specifications.

Image	Acquisition Date	Sensor Azimuth (deg)	Sensor Elevation (deg)
Graz North (left)	06/20/2007	218.6486	79.68948
Graz North (right)	06/20/2007	5.8984	64.55371

Table 3: Specifications of the stereo IKONOS scenes of Graz.

All presented software packages were used to orthorectify and derive the DSMs for both study areas. Depending on the software packages RF and/or physical models were used for orthorectification.

6 Orthorectification Results

6.1 QuickBird Data Nairobi - Geometric Correction Accuracy Analysis

QuickBird data is distributed in three different product levels: Basic Imagery, Standard Imagery and Orthorectified Imagery. The Nairobi scene was delivered in Level 2A, the standard imagery product corrected for systematic radiometric and geometric distortions. Three orthorectification methodologies were tested using PCI software:

- RPC orthorectification without any GCP
- RPC orthorectification with GCPs
- Rigorous physical model orthorectification with GCPs.

For the test with the GCP-based corrections eight GCPs and four GPS points as independent check points (ICPs) were used. The GCPs are well distributed in orientation and elevation over the dataset. The selected ICPs are rather situated in the central area of the scene. For the RPC-based orthorectification including GCPs two polynomial adjustment methods were tested. Figure 3 shows the distribution of the GCPs and ICPs. Table 4 shows a summary of the results with the root mean square errors of model computation over ICPs.

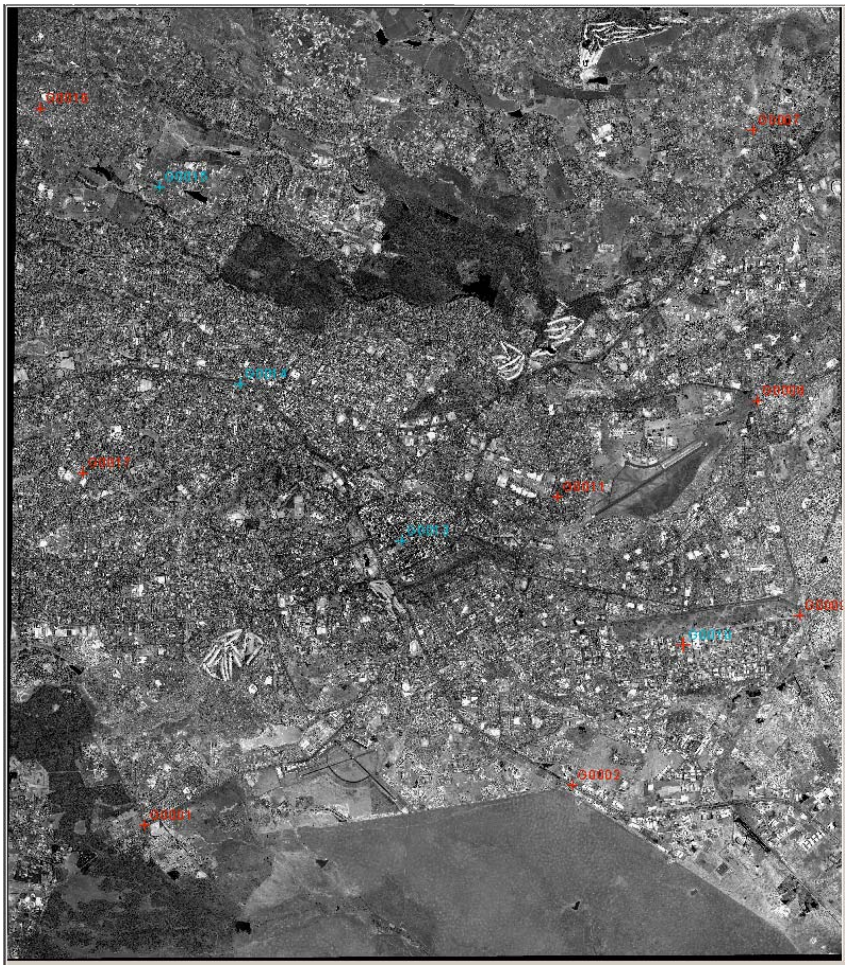


Figure 3: Distribution of selected GCPs (red) and ICPs (blue).

Model	No. of GCPs	No. of ICPs	ICP RMSE [m]	
			X	Y
RPC without GCP	0	4	17.32	29.96
RPC with GCP, 0-order	8	4	2.93	1.32
RPC with GCP, 1-order	8	4	0.83	1.41
Rigorous physical	8	4	0.31	0.45

Table 4: Comparison of RMS errors using three different orthorectification methods.

The table clearly shows that an orthorectification without any GCPs results in a weak result with large errors. If eight GCPs are used to adjust the rational polynomials functions a substantial improvement of geometric accuracy can be achieved. However, the best results are obtained with the rigorous physical model which requires a minimum of six GCPs. The RMSE is about 2.5 to 3 times lower than the RMSE obtained with the adjusted RPC orthorectification method. The 1-order polynomial adjustment resulted in a better orthorectification result than the 0-order polynomial adjustment.

It should be noted that the study area has only an elevation difference of approx. 300m. The error may rise for the RPC method, which is only an empirical/statistical model, for areas with rugged terrain because the complementary first order polynomial adjustment to the data may not be adequate enough to compensate for the errors.

The finding of this comparison clearly indicates that the rigorous model should be employed as a primary choice. The RF model is still a viable alternative when the accuracy requirement is not a high priority or when the number of GCPs is very limited. Similar analysis (Toutin et al., 2002) has confirmed our results. Moreover, it was observed that the RPC based orthorectification method is very instable depending on the terrain and the number and distribution of GCPs and is therefore not recommended (Toutin & Cheng, 2002).

6.2 IKONOS Data Graz - Geometric Correction Accuracy Analysis

As with data acquired by QuickBird satellite, IKONOS data is distributed in different product levels. The stereo IKONOS scene was delivered as standard geometrically corrected imagery product corrected for systematic radiometric and geometric distortions and is delivered together with an RPC-file for orthorectification. As for Nairobi and the QuickBird scene three orthorectification methodologies were tested for Graz and the IKONOS scene using PCI software:

- RPC orthorectification without any GCP
- RPC orthorectification with GCPs
- Rigorous physical model orthorectification with GCPs.

The models' accuracies were assessed by selecting 6 ICPs and calculating their RMSE. For the calculation of the DSM and the orthorectification all GCP including the ICPs were used. The ICPs were only selected for the accuracy and stability testing of the different models. The GCPs are well distributed in orientation and elevation over the dataset. Figure 4 shows the distribution of the used GCPs and ICPs respectively.

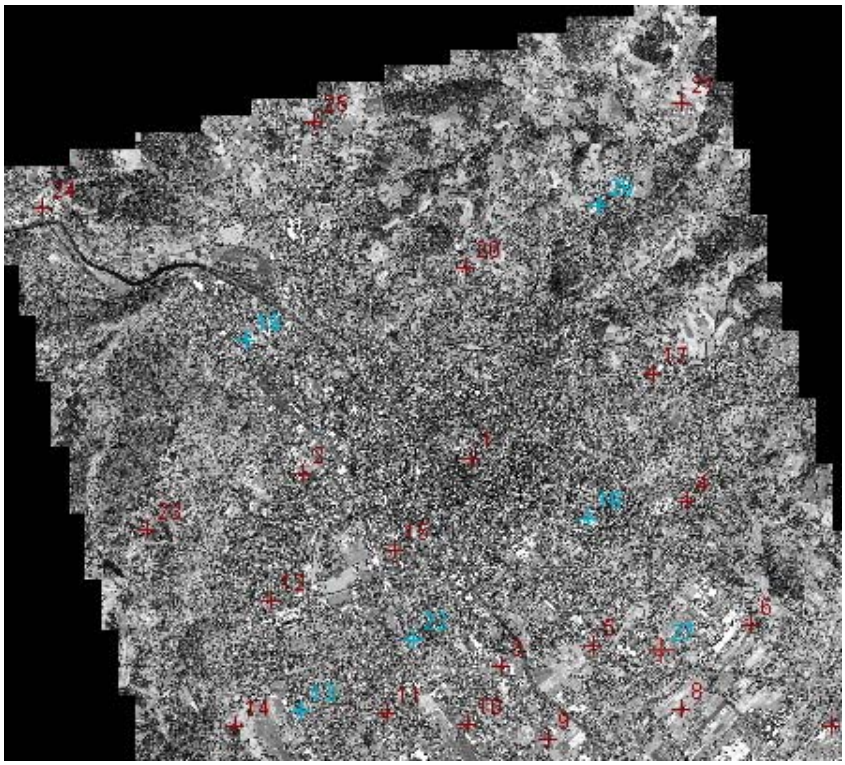


Figure 4: Overview of GCP and ICP distribution.

In all tested models the same GCPs and ICPs were used for the orthorectification and accuracy assessment. Table 5 shows a summary of the root mean square errors of the model computation over ICPs and the model GCPs.

Model	No. of GCPs	No. of ICPs	GCP Overall RMSE (using 26 GCPs) [m]		ICP RMSE [m]	
			left image	right image	X	Y
RPC without GCP (13 coefficients used)	0	6	0.32	0.55	1.56	2.68
RPC with GCP, 1-order	20	6	2.65 (max -10.33)	1.33	0.97	0.63
Rigorous physical	20	6	1.50	1.17	0.56	1.06

Table 5: Comparison of RMS errors using three different orthorectification methods.

The table shows impressively that the model with the lowest overall RMSE doesn't necessarily have the best position accuracy using unbiased validation with ICPs. The comparison of the RMSE for the RPC-based orthorectification using only the GPCs and the rigorous physical model shows clearly that the physical model achieves a better accuracy. Both orthorectifications were computed with the same GCPs and ICPs.

Looking at the RPC-based orthorectification using 1-order rational polynomial function one can see that the variability in RMSE is much larger looking at the maximum error of 10.33m for one of the GCPs in the left image. It indicates great accuracy variability and thus instability of the model and confirms the experiences of other studies (Toutin & Cheng, 2002). The RF model is still a viable alternative when the accuracy requirement is not a high priority or when the number of GCPs is very limited.

6.3 Discussion

The orthorectification tests with independent check points have shown that the rigorous physical model achieved the best accuracy. The physical models proved to be stable for both IKONOS and QuickBird data using few GPCs, eight for Nairobi, or a large number of GCPs, 26 for Graz. Consequently it is recommended to use the rigorous physical model for orthorectification of VHR satellite data if GCPs are available and a high geometric accuracy of the data is necessary. The RF model is still a viable alternative when high accuracy GCPs are very limited or not available. If the topography is rather flat in a dataset a 0-order polynomial adjustment with RF model orthorectification might be sufficient but a 1-order polynomial adjustment should be preferred if the terrain is rugged.

7 DSM Extraction Results

7.1 Digital Surface Model Comparison - Nairobi

The comparison analysis was done with four different software packages offering a DSM generation module. The software packages, their technical characteristics and manual interaction possibilities are described in detail in the theory chapter.

The stereo IKONOS scene was orthorectified using 9 GCPs with each of the software packages. For orthorectification either a physical model (PCI) or a RF model (ENVI, RSG, LPS) including nine very accurate GCPs was used. The parameter settings vary slightly for all generated DSMs due to different possibilities of parameter setting for every software package. They are listed in Table 6.

Software Settings	PCI	ENVI	RSG	LPS
No. of DSMs generated	1	3	2	2
Resolution [m]	1	1	1	2
Extraction Detail	"high"			
Number of scans		maximum		
Topographic Type				mountain
Object Type				high-urban
Terrain Relief Type	"high"			
Adaptive Changes				Yes/No
Coefficient Limit		0.7		0.8
Matching Window Size				7x7 (default)
Search Window Size		5x5, 9x9, 13x13	13x47, 6x23	27x3

Table 6: Overview of the parameter settings selected for PCI, ENVI, RSG and LPS software packages.

PCI offers only few parameters to be adjusted. Due to large view angle differences and tall buildings in the two scenes the most extreme settings for “extraction detail” and “terrain relief type” were selected.

In ENVI additionally the “number of scans”, “correlation coefficient limit” and the “search window size” can be adjusted. The maximum number of scans was selected, as well as the by ENVI suggested correlation coefficient limit. Three different search window sizes were tested.

LPS offers a larger variety of parameters to be set. Again an extreme setting for “topographic type” as well as the respective “object type” was selected. As correlation coefficient the by LPS suggested limit was selected. The matching window size was set to the default value and the search window size was set to the maximum, which is defined by the selected “high-urban” object type setting. LPS additionally offers a parameter called “adaptive changes”. If the corresponding checkbox is enabled the search window size, correlation window size, and correlation coefficient limit may be adjusted automatically. LPS computes and analyzes the terrain features after each pyramid and sets the strategy parameters accordingly.

With RSG the search window size was set to the calculated pixel and line differences between the two scenes for the tallest building found in the test area. Additionally, a search window half this defined size was tested to see what the differences and consequences for the two DSMs generated with RSG are. RSG performs the matching not only on the two grey-value images itself. It additionally offers a variety of features (e.g. textural features) to be calculated and matched. A stack of features offered by the software package which is suitable for IKONOS stereo data was selected.

After the generation of all DSMs several qualitative and quantitative accuracy comparisons were performed because no reference surface model or building height measurements are available for the Nairobi test area. The quantitative analyses were done with external sources of unknown origin and accuracy. Moreover ground elevation measurements taken from the measured differential GPS points were analysed. The following comparisons were conducted:

- GPS-based ground elevation comparison
- High-rise building heights comparison
- Building heights (Canty’s methodology) comparison
- Vertical profile analysis
- Visual evaluation.

7.1.1 GPS-based ground elevation comparison

First, the available differentially measured GPS heights present in the study area were compared with the generated DSM heights. The comparison was done for the original datasets of 1m and 2m respectively. Additionally, the DSMs were resampled to 5m, 10m, 20m, 50m, and 90m and cross-compared to see how the generated elevation values change depending on the geometric resolution. Moreover, the 90m resampled DSMs were compared with the SRTM heights. Six GPS points were compared. The results for this comparison are listed in Table 7 and Figure 5.

Point ID	Elev. asl.	PCI	RSG 13x47	RSG 6x23	ENVI 5x5	LPS adaptive	LPS default	ENVI 9x9	ENVI 13x13
G0002	1632.21	5.21	7.33	7.56	1.42	1.67	1.69	4.67	4.41
G0003	1616.43	3.43	7.96	8.02	0.60	2.95	3.09	5.32	5.14
G0009	1615.09	7.09	11.01	11.70	3.56	5.36	5.24	4.83	4.83
G0010	1616.70	1.70	3.40	4.81	0.64	-0.66	-0.87	2.56	2.45
G0011	1630.40	-0.60	2.38	2.34	0.59	0.51	0.54	2.03	2.11
G0013	1650.35	5.35	-6.05	-1.73	0.47	2.40	2.64	-0.20	2.16
RMSE		4.92	7.64	7.62	1.79	3.05	3.09	4.10	4.11

Table 7: Height differences between the differentially measured ground elevations and the elevations extracted from the generated DSMs.

In the subset only six GPS elevation points were available for the comparison with the calculated DSMs. The best elevation accuracies were achieved with the ENVI 5x5 DSM. The largest RMSEs were calculated for the DMSs generated with RSG. The RMSE for both LPS DSMs are similar as well as the PCI-, ENVI 9x9- and ENVI 11x11-DSMs. The rather large differences for Point G0009 could be due to its position at the beginning of a bridge over a railway ditch at a much lower elevation.

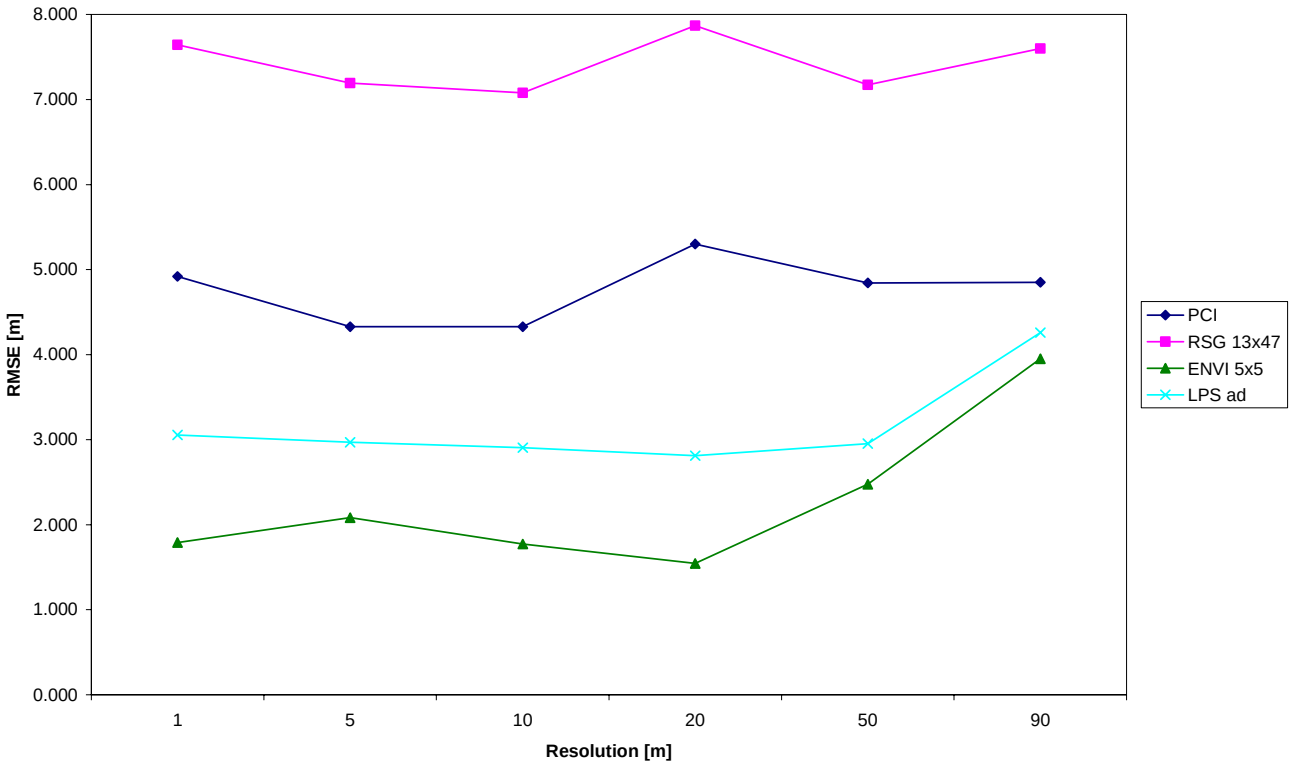


Figure 5: Comparison of the scaled DSM heights with the GPS heights (1m – 50m) and the SRTM heights (90m).

In general it can be said that the RMSE doesn't change much with scaling, however PCI, RSG and ENVI get slightly better at 5m scaling. It could be that erroneous height influences due to noise or a positioning error are eliminated.
Comparing the calculated DSM heights at 90m resolution with SRTM shows that the differences are larger than compared with the measured GPS elevations. This is due to the difference between the actual elevation above sea level and the SRTM data measuring the surface height (including vegetation and buildings) which results in

a higher value for every 90m pixel whereas the generated DSMs tend to underestimate the actual elevation asl. The RMSE for the SRTM heights compared with the 16 GPS points is 4.169m and shows that SRTM has a similar error as the generated DSMs if compared to ground elevation measurements.

7.1.2 High-rise Building Heights Comparison

Since we were not able to measure building heights in Nairobi for the evaluation of the DSMs the internet was searched for high-rise building databases. For the Nairobi test area building heights for 21 (22 and 19 respectively) prominent high-rise or historical buildings with heights between 30m and 140m were found and compared to the heights of the generated DSMs (http://www.emporis.com/en/wm/ci/bu/?id=100051, visit: 20.11.2007). A geographical overview of the buildings is given in Figure 6. The DSMs generated with LPS are slightly smaller in size and do not cover four buildings at the border of the calculated DSMs. The number of buildings is accounted for in the calculation of the RMSE.

The results of the comparison are listed in Table 8. The tall buildings were clearly underestimated by all software packages. The DSM generated with LPS software applying the “adaptive” parameter setting achieved the lowest RMSE of 40.85m whereas ENVI created the DSM with the highest RMSE of 64.48m. Looking at the different parameter setting selections chosen for the four software packages it can be seen that the different search window sizes chosen with RSG software had a high impact on the building height estimation. The RMSE with the large search window achieved an RMSE of 48.87m and the smaller search window an RMSE of 57.44m. The large search window size was defined by counting the matching distance between the two stereo images for the largest building (roof-top) in the CBD. If a smaller search window size is defined the software doesn’t find the corresponding pixel in the search window for large buildings which results in a smaller, wrong matching distance and consequently in a lower building height estimation.

For the three matching window size scenarios chosen in ENVI, the software achieved the best RMSE for the smallest matching window size. This makes sense since the chance for mismatches between two corresponding pixels is reduced with a smaller matching window. It is recommended to choose a smaller matching window sizes for areas containing large degrees of topographic relief, grey level variation, and colour intensity variation.

Looking more closely at the height differences of the different buildings (Table 8 and Figure 7) it can be seen that smaller buildings with heights of maximum 50m were calculated well by LPS, PCI and RSG; differences were often <10m. Buildings higher than 50m are often not detected at all or estimated with a larger error. In particular the “New Central Bank Tower”, the largest building with 140m of height was not detected by any of the software packages. The “Kenyatta Conference Center” and the “Cooperative Bank House” 105m and 83m high were detected well by LPS software using the “adaptive” parameter set as well as the “Posta House”, 81m high, detected well by PCI.

Figure 6: Geographical overview (2km by 2km) of the compared high-rise buildings in the Central Business District (CBD) of Nairobi.

ID	Building Name	Height [m]	Floors	∆h PCI	∆h RSG 13x47	∆h RSG 6x23	∆h ENVI 5x5	∆h ENVI 9x9	∆h ENVI 13x13	∆h LPS def	∆h LPS ad
1	New Central Bank Tower	140	?	107.67	119.69	122.02	132.52	136.30	136.19	130.98	130.97
3	Kenyatta Conference Center	105	33	111.97	83.92	91.66	102.12	99.86	106.93	91.24	7.97
8	Cooperative Bank House	82.9	25	82.03	67.09	70.26	77.64	75.63	78.32	19.67	0.34
10	Posta House	81	27	1.30		86.41	80.27	82.04	81.16		
11	Nyayo House	80	26	88.87	80.00	71.79	75.36	78.48	79.34	41.86	21.21
12	Anniversary Towers	80	26	49.30	64.28	89.60	69.69	69.88	75.74		
18	International House	66.1	17	41.10	33.96	37.56	64.23	65.86	64.24	8.57	6.41
19	AmBank House	66	22	54.13	43.39	79.89	60.38	66.51	65.50		
21	View Park	60	20	11.60	44.05	64.57	51.44	51.67	54.88		
24	City Hall	60	18	40.77	40.36	39.01	44.39	45.25	45.30	46.01	45.41
27	Corner House	54	18	17.47	35.14	36.94	46.65	43.59	43.50	42.87	44.15
28	Nation Centre	51	17	14.20	26.90	29.75	52.02	39.65	45.23	18.31	15.84
21	Nairobi Hilton	61	20	35.60	41.24	43.45	46.23	50.99	58.14	47.28	82.78
30	Treasury House	47.9	15	49.73	-11.98	43.16	43.43	45.87	48.06	-2.30	-2.12
32	Hotel 680	46.9	14	-5.83	31.60	28.20	39.79	43.09	41.82	5.80	5.67
33	Bima House	45.1	13	13.87	-2.10	21.24	39.32	40.99	40.94	-4.49	-3.55
34	Nairobi Safari Club Hotel	45	15	-0.17	13.26	11.83	41.43	31.59	45.52	-3.85	-4.60
35	Barclay Plaza	45	15	48.57	36.59	36.10	35.35	47.64	46.52	-2.76	-17.23
36	Office of the President	43	14	23.73	18.37	17.02	30.80	31.81	36.20	0.99	0.99
41	Development House, West Tower	39	12	2.23	-4.83	14.19	36.90	38.04	42.33	3.65	3.64
43	IPS Building	36	12	16.63	11.22	7.55	18.43	23.50	24.83	-5.22	-5.04
44	Harambee Plaza	39.9	12	1.13	17.36	30.92	37.68	35.38	38.60	26.90	24.98
45	Grand Regency Hotel	30	10	7.97	-5.77	-5.18	29.87	21.95	27.55	-10.35	-13.36
	RMSE (Root Mean Square Error)			50.16	48.87	57.44	61.55	62.25	64.48	44.21	40.85
	n			23	22	23	23	23	23	19	19

Table 8: Comparison between prominent high-rise or historical buildings and the generated DSM-derived building heights in Nairobi.

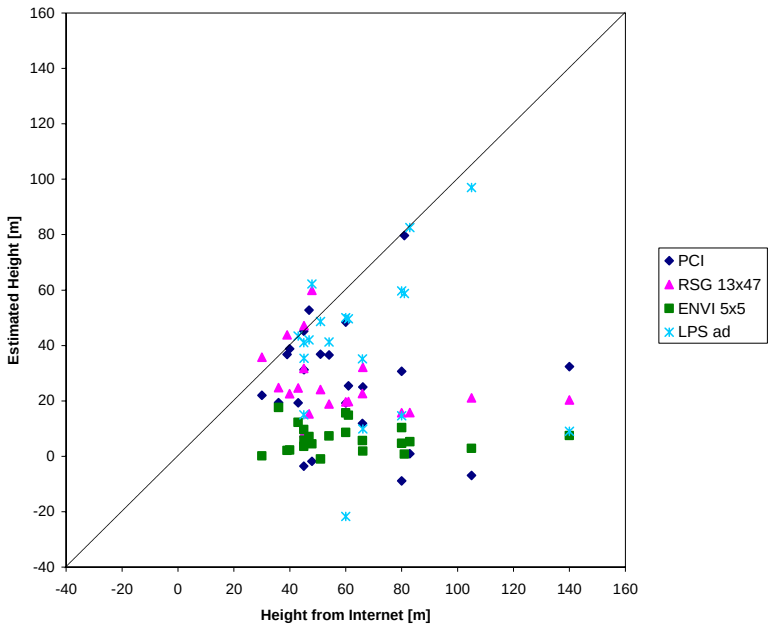


Figure 7: Calculated building heights vs. building heights known from internet sources. The best scenarios for all software packages are shown for the Nairobi test area.

7.1.3 Building Heights Comparison (using RF model approximations as reference heights)

In a previous chapter it was discussed that data providers are delivering VHR satellite data with RPCs in order to let the user orthorectify the satellite data with RF models. The RF model is capable of representing the camera model well and can be used as a replacement for it. Canty (2007) has developed a tool, reading every stereo scene with its corresponding RF model which can be used to determine object heights. To do so the bottom and top of the object has to be determined and an elevation height entered manually. The determination of the bottom and top of medium sized and small buildings is not easy. Depending on the image acquisition angel and the sun azimuth angle often building edges are not clearly visible. Additionally, the two points to be determined are close together for small buildings. A slight change in setting the point (one or two pixels) can result in a varying building height difference of several meters. Still, it was tried in this analysis to determine the height of a number of different building types which edges are well visible. The determined heights were compared with the heights calculated by the software packages and the results summarized in Table 9.

The RMSE shows that LPS (both scenarios) and RSG 13x47 are estimating the heights best. The mean absolute error however favour PCI and LPS DEF, followed by RSG 13x47. The absolute sum is lowest for PCI.

Analyzing the comparison by building type shows that the smallest buildings, the shacks are not well detected by the generated software packages. All software packages are underestimating the height of the two shacks. It could also be said that the height of the shacks are not detected at all. However, a visual inspection shows that the structure of the buildings and narrow roads is detected by some of the software packages, such as RSG, ENVI and PCI. Only RSG is estimating well one of the two shacks. The two medium-sized industrial buildings as well as the multi-storey buildings are under- and overestimated but overall their heights are estimated well. Again it can be seen that the larger buildings are more under- or overestimated than the medium-sized buildings.

Since the reference heights of this analysis based on the satellite data and the delivered RPCs they are not completely independent and thus these results should be interpreted with care. Also the selection of the building edges at the ground and the roof of a building is highly unstable and influencing the calculated height of the building. It can easily be calculated with height errors of several meters if only an error of one pixel is done during the edge and ground point definition.

Building Type	Height [m]	Δh PCI	Δh RSG 13x47	Δh RSG 6x23	Δh ENVI 5x5	Δh ENVI 9x9	Δh ENVI 13x13	Δh LPS AD	Δh LPS DEF
Storage Building	22	5.00	10.07	17.53	21.53	22.45	20.81	16.03	14.60
Multi-Storey Building 1	10	1.00	-0.85	-3.63	4.85	-1.40	4.39	-1.06	-1.30
Multi-Storey Building 2	14	1.00	-9.76	-9.71	9.79	0.50	10.28	0.99	0.45
Shack 1	3	3.00	4.84	4.88	3.65	2.81	3.16	3.16	3.64
Shack 2	3	2.00	1.05	0.08	2.40	2.82	3.23	2.61	2.45
Industrial Building 1	10	0.00	-3.08	-3.03	1.26	0.88	5.16	3.16	3.14
Oil tank 1	14	6.00	0.50	-4.27	6.30	12.43	13.82	9.45	12.28
Oil tank 2	19	10.00	0.25	4.03	2.87	18.92	19.85	7.09	6.71
Oil tank 3	16	-3.00	8.52	13.01	-0.12	14.78	17.92	12.39	10.76
Hangar	30	14.00	9.70	9.99	14.11	17.30	17.71	14.87	14.52
Office Building	29	-4.00	-15.58	-14.42	-0.16	15.52	21.88	2.76	1.95
Industrial Building 2	7	-2.00	-4.03	-3.99	-0.71	-0.45	-0.78	-0.33	-0.38
Multi-Storey Residential	19	1.00	-3.99	-3.55	14.46	15.74	18.80	9.63	7.82
High-Rise Building Commercial	33	29.00	19.02	21.09	32.86	29.24	32.66	13.19	8.97
RMSE		9.78	8.92	10.50	12.83	14.93	16.93	9.09	8.37
ABSOLUTE SUM		81.00	91.24	113.20	115.08	155.23	190.45	96.72	88.95
MEAN ABSOLUTE ERROR		6.23	7.02	8.71	8.85	11.94	14.65	7.44	6.84

Table 9: Comparison between calculated building heights by RF model (reference) and the generated DSMs.

7.1.4 Vertical Profile Analysis

In order to understand how well the buildings are extracted several vertical profiles of different building types and densities were analysed for all generated DSMs. The profiles show how well buildings, their shapes and heights were extracted. Four representative profiles are described and discussed more in detail. A theoretical model of the buildings' shape and size was inserted in the chart in order to better visualize the differences between the generated DSMs and the theoretical building model.

Since the LPS scenarios resulted in very similar profiles only one of them is shown. Due to the low performance of the RSG 6x23 scenario the analysis focussed on the 13x47 scenario. For ENVI the scenario 5x5 was selected because it performed best among the three ENVI scenarios. For the following four building types a profile was set:

- a high-rise building
- an industrial building
- a residential area
- an illegal settlement area.

The heights of high-rise buildings are often underestimated due to large matching distance which increases the chance for mismatching and the calculation of wrong building heights. Additionally, occlusion and shadow areas complicate the matching process. However, acceptable results were achieved for buildings not taller than 50m.

In the first profile a building of 45m was selected.

Figure 8 shows the overview image of the profile line as well as the profiles extracted from the generated DSMs. Additionally, a model profile was introduced (blue line) to be able to compare the generated results with an ideal building profile.

The building was detected by all software packages except for ENVI. The height was well estimated by LPS and PCI which over- respectively underestimated the height by approximately 2.5m. RSG had problems extracting the building and did underestimate the height by approximately 12m. ENVI failed completely in extracting the building and interpolated the area. All software packages failed to generate the square shape of the building, especially the building wall facing the South was not calculated as a straight, vertical line indicating a large, abrupt height difference.

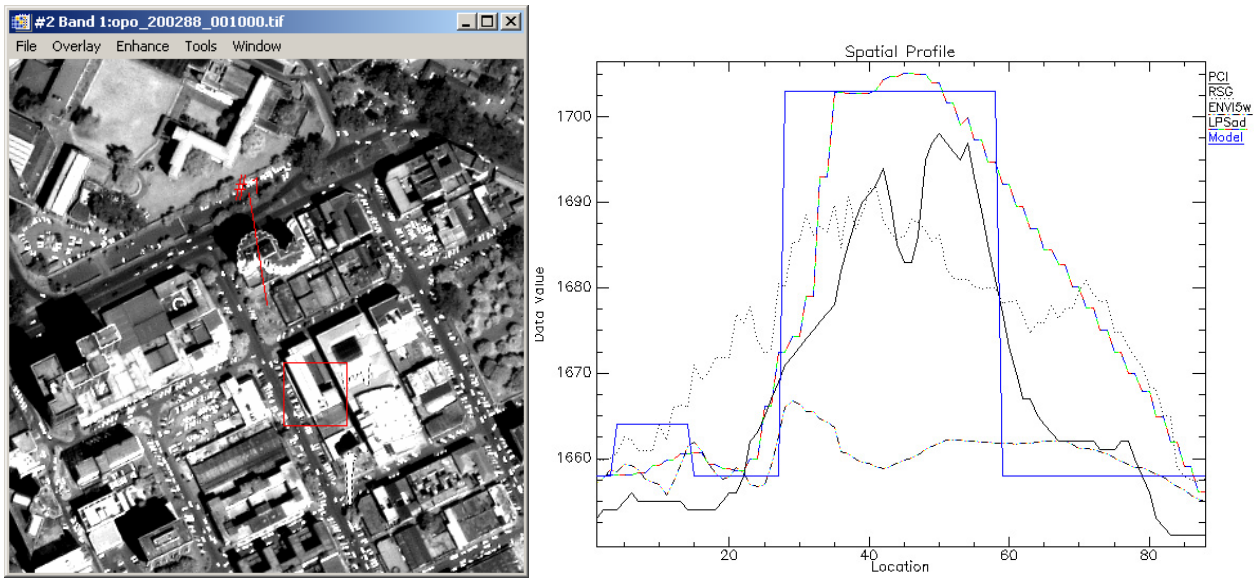


Figure 8: Profile line overview and profiles for a high-rise building.

The second profile represents a large industrial building. Its roof has five and a half small saddles. Next to the building a large tree is standing which should result in six peaks in the profiles. The building is estimated to be approximately 10m to 15m high including the roof saddles of approx. 4m of size. In Figure 9 the blue line is representing an approximation of an ideal profile of the building.

PCI, RSG and ENVI detected the six peaks and the building itself. However, the height and intervals differ. LPS didn't detect the building at all. For all DSMs the ground elevation differs as well as the building size which was extracted with a height between 10m and 14m. The saddle roof height was estimated to 3.5m to 4m. Overall it can be said that the structure of the roof is detected as well as the building as such.

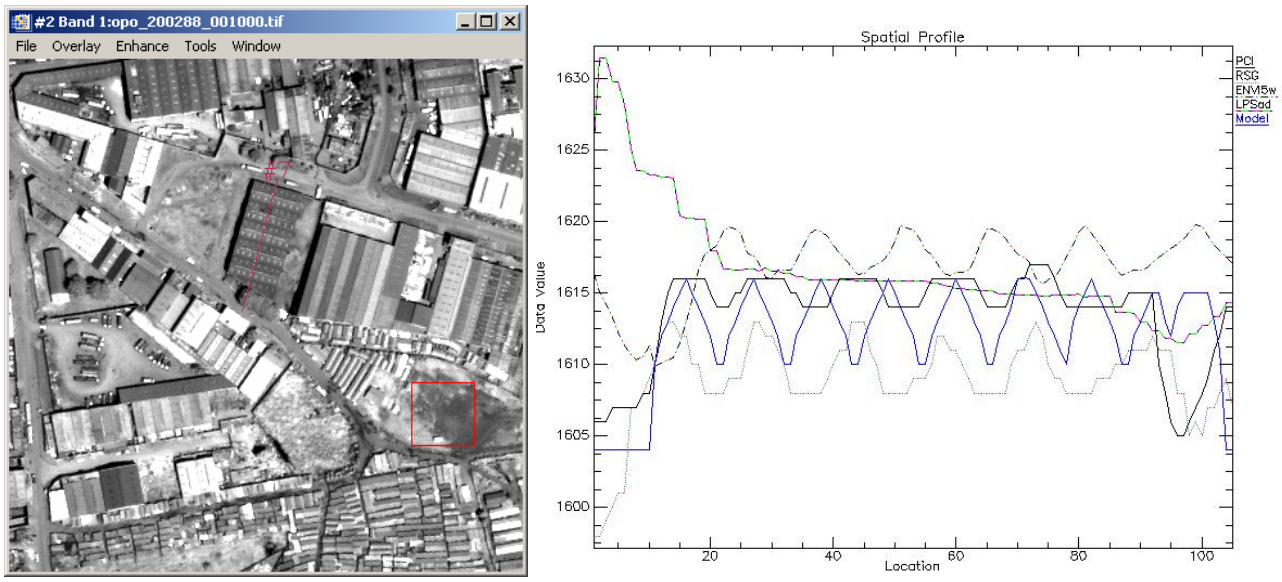


Figure 9: Profile line overview and profiles for an industrial building.

The next profile (see Figure 10) was laid in a residential area. The section contains two row houses with a lawn and maybe a shed in between. The houses are estimated to be approximately 4m high.

They were not well detected by any of the software packages. ENVI detected both houses but also an additional peak at the end of the transect. The height was estimated to 3m-4m. PCI detected the shed in between and the second house but not the first one. The height was estimated at about 2.5m. RSG extracted a lot of objects at a very low elevation level which can not be clearly assigned to any of the buildings and seems to be noise. LPS completely failed to detect any of the buildings.

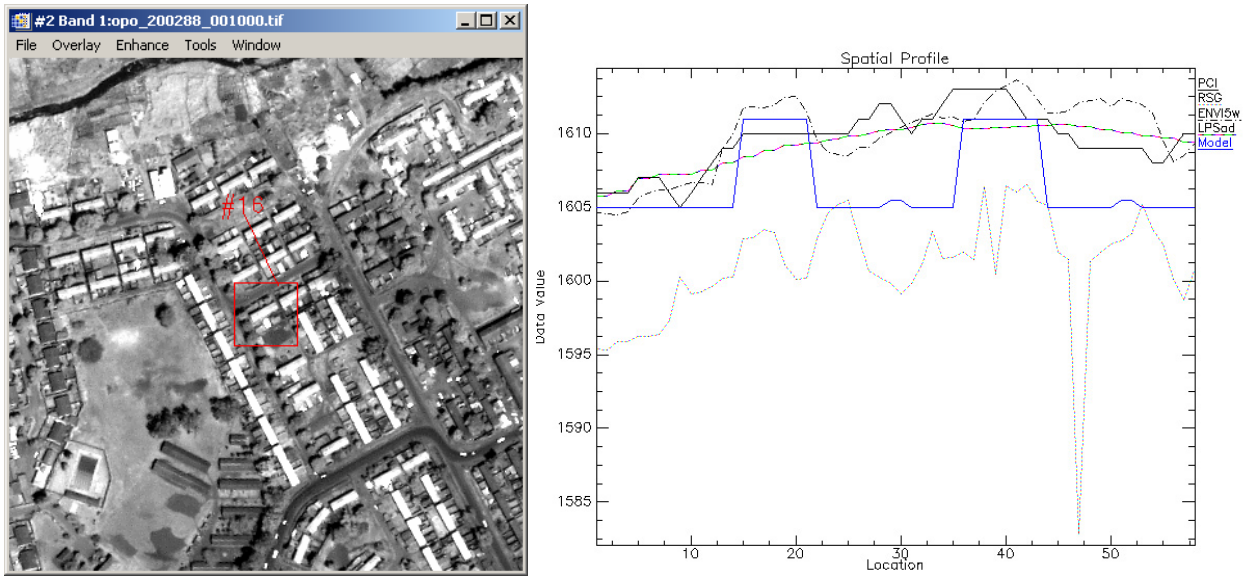


Figure 10: Profile line overview and profiles for residential row houses.

Finally, profiles of an informal settlement were analysed. The profile line was laid across five buildings. The buildings of informal settlements are typically small in size and height and stand close together. They are difficult to be extracted due to noise which is estimated in the range of one floor or 3m to 4m respectively. Figure 11 shows that ENVI and PCI were succeeding in extracting three of the five buildings. RSG extracted three objects at a low elevation level that are not related at all in shape and size to the five objects to be extracted. LPS extracted one object only. This example shows that the software packages have problems extracting small buildings of 3m-4m of height. However, the visual analysis shows that the buildings structures are correctly extracted by ENVI 5x5 (Figure 15) and partially by RSG.

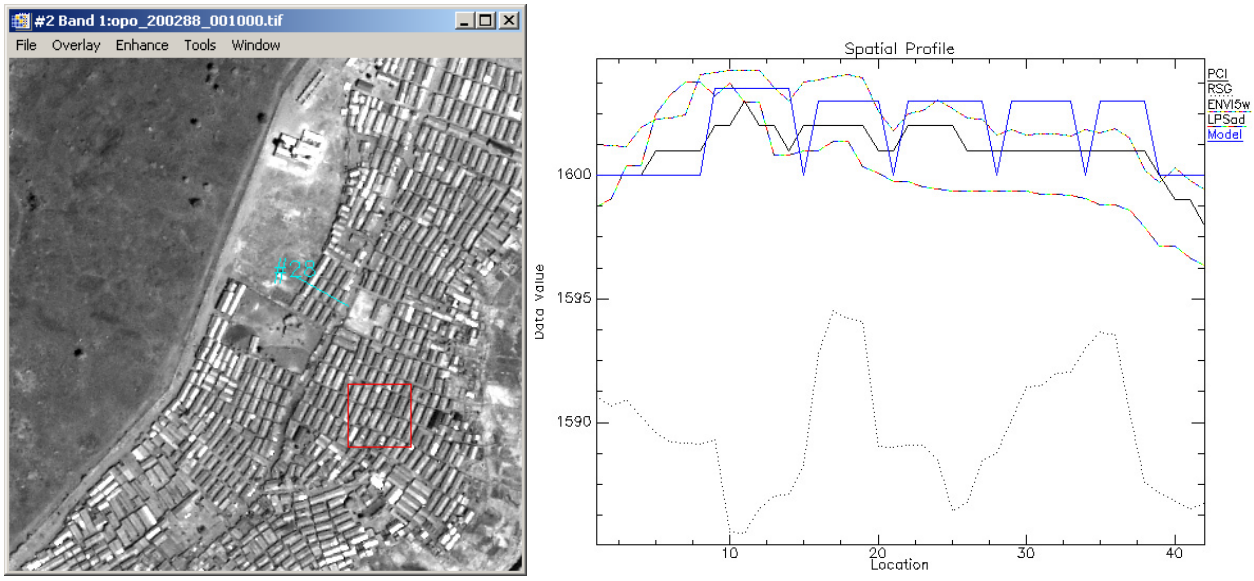


Figure 11: Profile line overview and profiles for an illegal settlement.

7.1.5 Visual Evaluation

The visual evaluation was done by selecting four subsets of Nairobi representing different building types:

- The central business district with large high-rise buildings and a high building density
- industrial area of Nairobi with mostly rectangular large, medium-size buildings
- residential area with small, multi-apartment row-houses and single houses
- illegal settlements with small, rectangular buildings and narrow roads/paths in between.

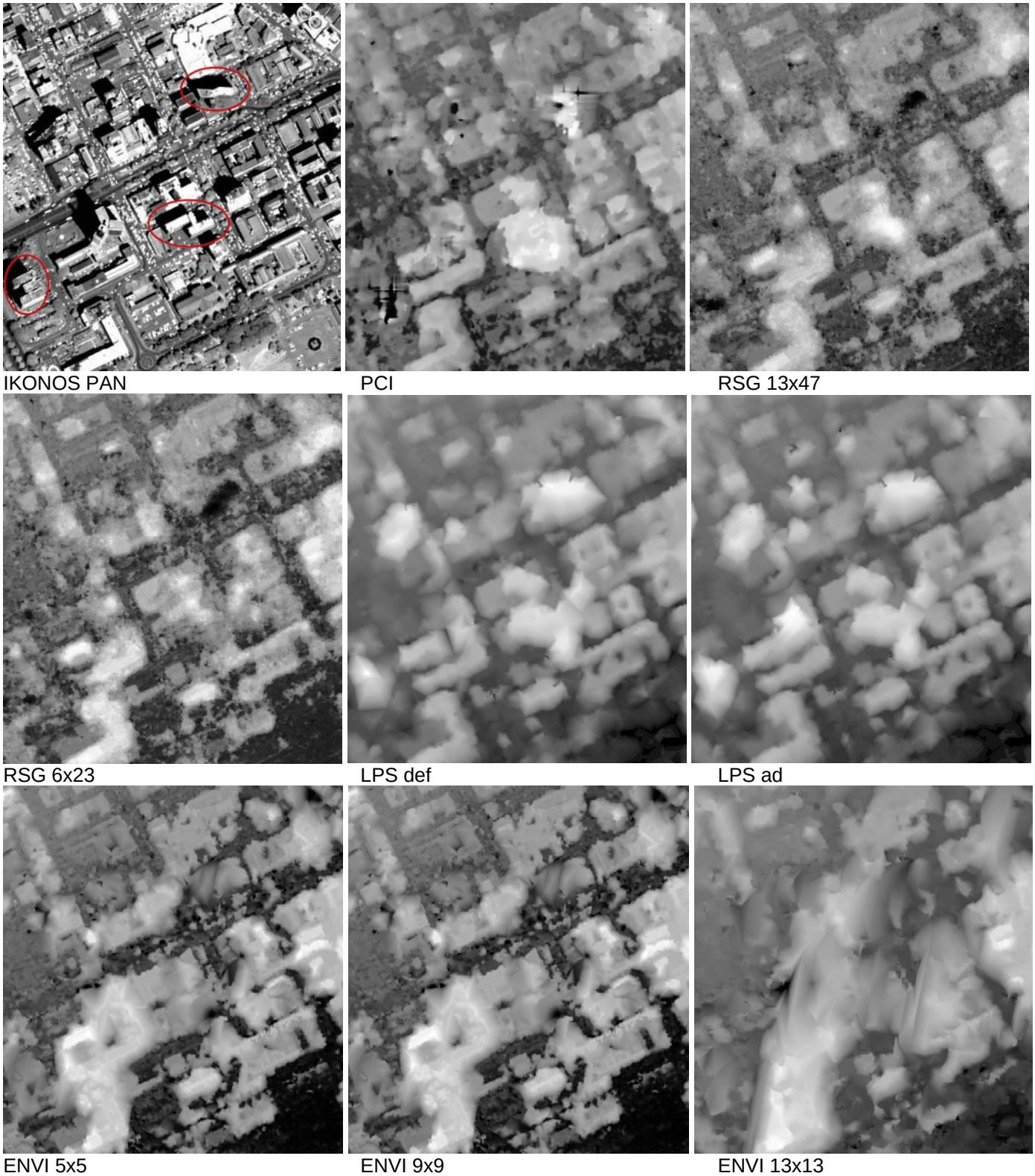


Figure 12: Visual comparison of all DSMs. Large buildings causing mismatching in the CBD are circled.

In a DSM buildings should be represented by their clear outline and their accurate height. Visual inspection (Figure 12) shows that for the CBD this was done best by PCI and RSG. Most buildings are well represented. However, the circled, rather complex and tall buildings caused problems for all software packages due to large distortion and shadow areas which results in a lack of corresponding pixels in one of the stereo pair images and thus in erroneous height extraction. The DSMs generated with LPS and ENVI show a lot of interpolated blurry areas where the outlines and heights of these tall building complexes should be. The three different ENVI scenarios show that the larger the selected search window size the more severe the interpolation errors. LPS created two rather blurry DSMs containing little details.

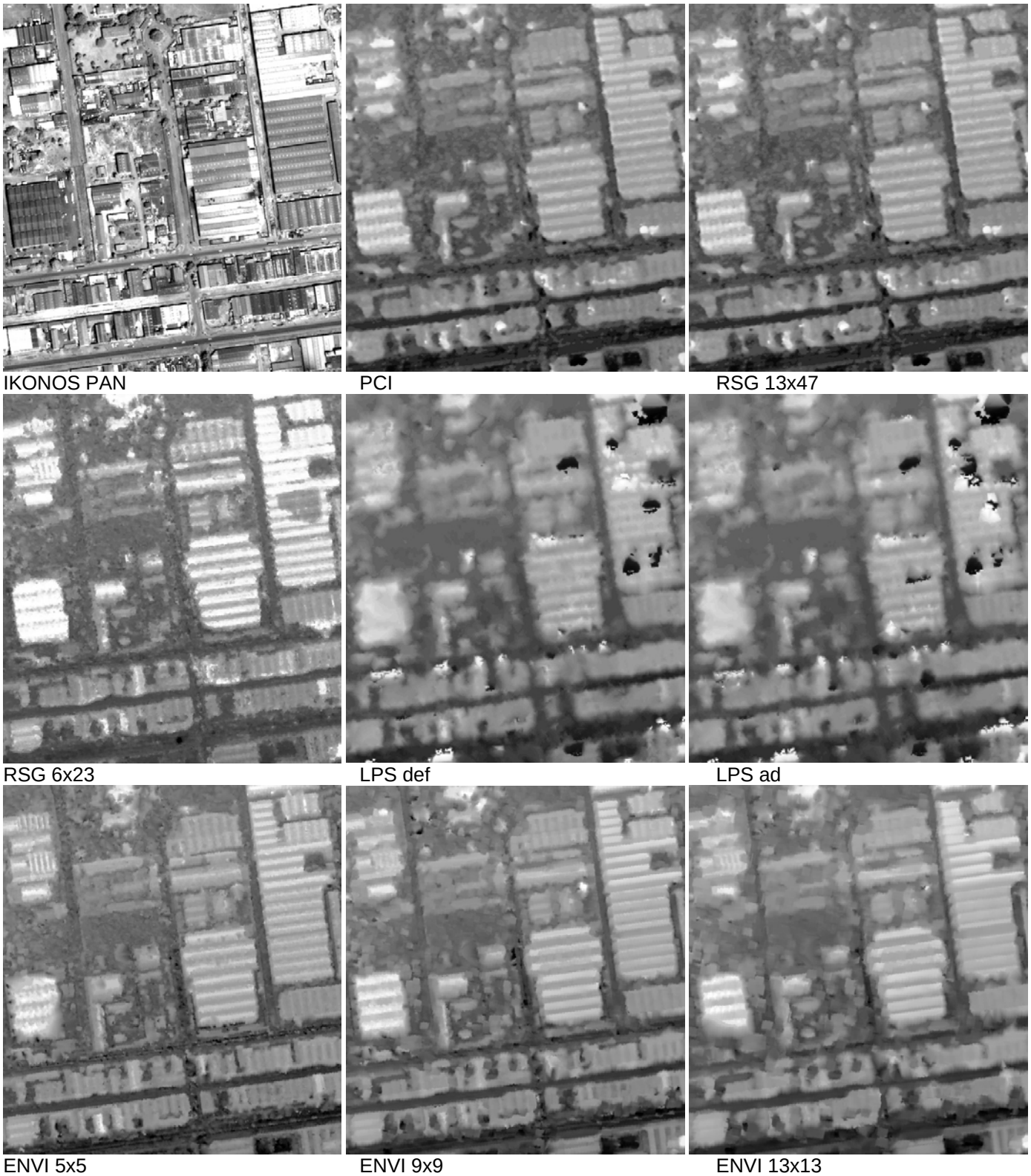


Figure 13: Visual comparison of all DSM scenarios representing an industrial area in Nairobi.

The large industrial buildings are represented well by PCI, RSG and ENVI (Figure 13). The shapes of the roofs were generated with detail. Only few erroneous artefacts were created. The best result was achieved with ENVI 5x5. LPS again created blurry DSMs containing a number of erroneous peaks and sinks. The scenarios ENVI 9x9 and 13x13 created blurry DSMs as well.

The residential subset (Figure 14) shows that both RSG scenarios as well as ENVI 5x5 were able to detect the small structures of the row houses and single buildings in the area. However, both RSG scenarios contain some artefacts due to trees. PCI achieved a slightly worse result followed by the blurry results of ENVI 9x9 and ENVI 13x13 and LPS where almost no details were detected.

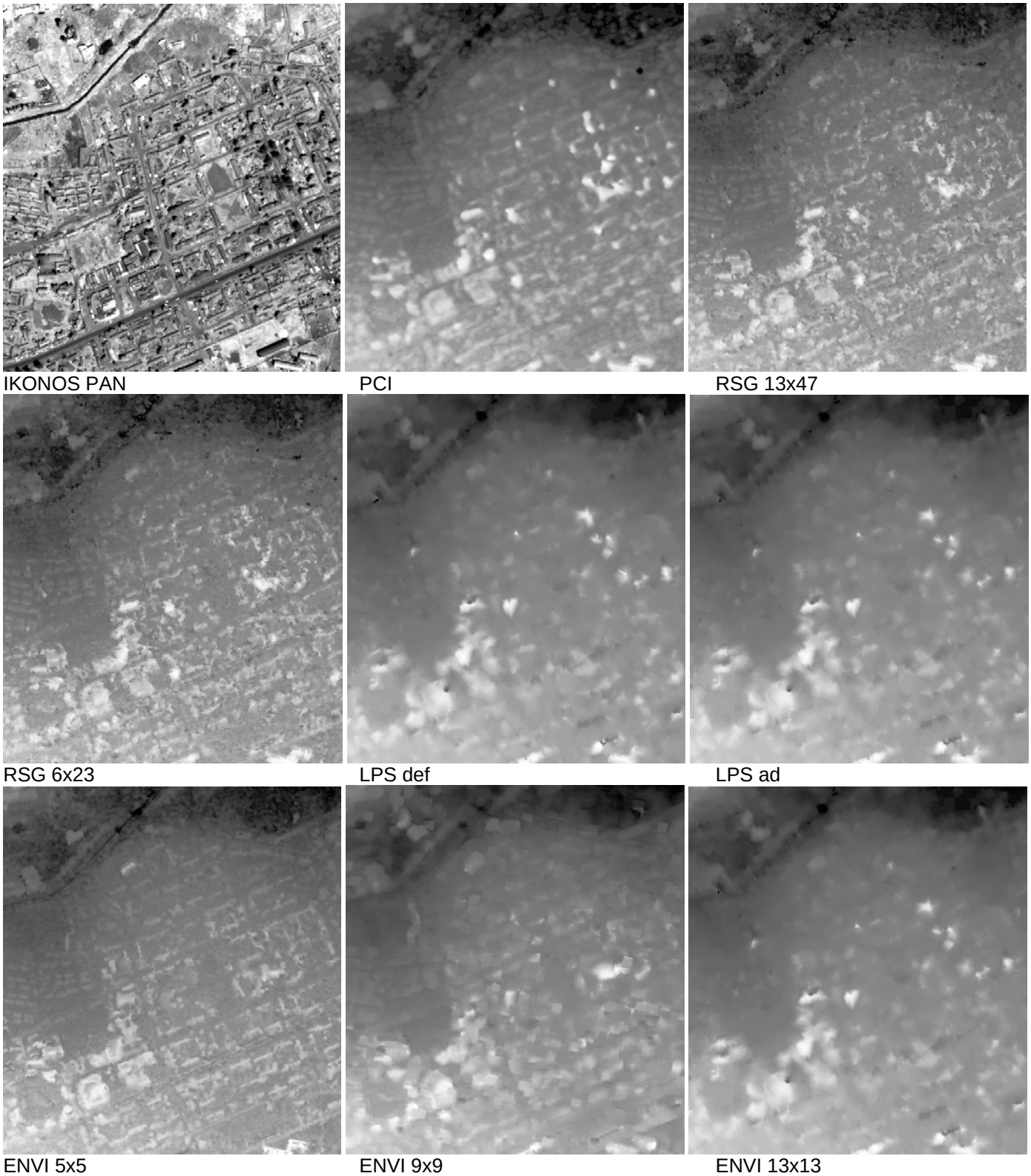


Figure 14: Visual comparison of all DSM scenarios representing a residential area (row houses).

The last subset (Figure 15) is an area representing illegal settlements consisting of small shacks with an approximate height of 3 meters. Due to their small size and density these buildings are very difficult to detect. RSG and ENVI 5x5 generated the best visual results detecting not only the roads in between the buildings but also identifying the single shacks if they are not standing too close to each other. PCI generated a DSM containing the main roads and the wider paths but not the single buildings. LPS didn't detect the small structures at all.

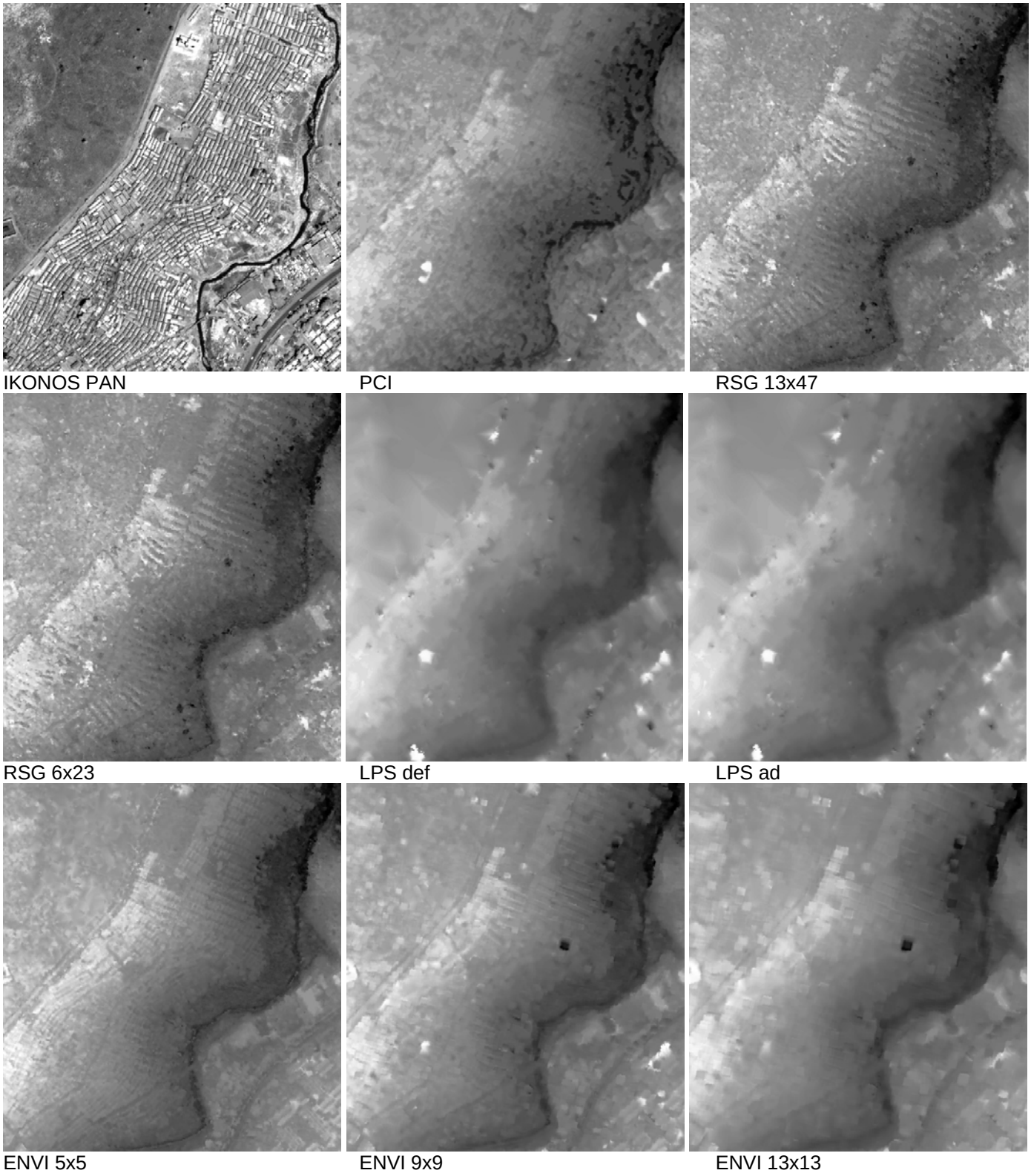


Figure 15: Visual comparison of all DSM scenarios representing an illegal settlement.

Summarizing the results it can be said that both scenarios of RSG achieved visually very good results and differences were small. PCI achieved good results for high-rise and industrial buildings but failed in extracting details of small and very small houses. ENVI achieved a detailed visual result with the 5x5 scenario. Except for high-rise buildings the houses were detected well with a high grade of detail. The 9x9 and 13x13 scenarios are less detailed and contain a high amount of interpolated areas. LPS failed in achieving an acceptable visual result. All DSMs are blurry and contain large interpolated areas; object structures are barely detected.

7.1.6 Discussion

Due to lack of reference data e.g. building heights and/or building outlines or a high-resolution digital surface model the generated DSMs were evaluated by comparing them with reference height data taken from the internet, the GPS ground elevation data collected in the field, and data calculated by an alternative height extraction methodology. Additionally, two qualitative tests were conducted to come to a conclusion in terms of DSM quality relating to building height and shape extraction.

The five evaluation tests have shown that all software packages created DSMs that performed well in at least one of the tests. They all have advantages and disadvantages. Table 10 gives an overview of the performances of all scenarios in the conducted analysis. The computation time is included in the table as an additional criterion.

Overview DSM Performance	PCI	RSG 13x47	RSG 6x23	ENVI 5x5	ENVI 9x9	ENVI 13x13	LPS AD	LPS DEF
High-rise building height comparison	**	*	--	--	--	--	***	***
Building heights (Canty's methodology) comparison	**	***	*	*	*	-	*	*
Vertical profile analysis	**	***	no data	**	*	--	-	-
GPS-based ground elevation comparison	-	-	-	***	*	*	**	**
Visual evaluation	***	***	**	***	--	--	-	-
Computation time	***	--	--	***	***	***	***	***

Table 10: Overview of DSM performance for all tests. Classification: *= very good, **= good, *= ok, -= bad, - - = not acceptable.**

Height accuracy as well as clear building shape extraction is of great importance for the use of DSMs in information extraction for settlement analysis and mapping. The highlighted tests are representing these criteria best. Judging them it can be concluded that overall the PCI and RSG 13x47 scenarios performed best. The two software packages PCI and RSG should be favoured for DSM extraction. However, a big disadvantage of RSG is the computation time. The calculation of a DEM of one stereo IKONOS scene took approximately one week whereas the other software packages took one day to extract the DSM. PCI performed well in all relevant tests. It didn't fail completely in any of the important criteria. ENVI 5x5 performed very well in the visual evaluation and the vertical profile analysis. However, it completely failed in correct building height estimations which is essential in settlement information extraction. LPS created large interpolation areas and failed in detailed building shape extraction which is another essential criterion to be fulfilled. Surprisingly LPS performed well in building height extraction but considering its weaknesses it is not suitable for urban DSM extraction. Consequently, PCI and RSG are recommended for urban DSM extraction.

7.2 Reference-Based Digital Surface Model Comparison - Graz

In the Graz test case again the four software packages ENVI, PCI Geomatica, RSG and LPS were tested. The selected parameter settings for all generated DSMs and the specific software packages are listed in Table 11.

Software Settings	PCI	ENVI	RSG	LPS
No. of DSMs generated	1	2	1	1
Resolution [m]	1	1	1	1
Epipolar Projection		automatic/calculated		
Extraction Detail	"high"			
Number of scans		maximum		
Topographic Type				Rolling Hills
Object Type (Defined Areas)				High Urban, Low Urban, Open Area
Terrain Relief Type	"high"	moderate		
Adaptive Changes				No
Coefficient Limit		0.7		0.8
Matching Window Size				7x7
Search Window Size		5x5	18x54	21x3, 19x3, 11x3

Table 11: Overview over the parameter settings selected for PCI, ENVI and RSG software packages.

Following the experiences in the Nairobi test case the settings for the DSM generation were selected similarly. For PCI the settings remained the same as in the Nairobi test case. For ENVI the best scenario in the Nairobi test case was selected and for RSG the search window size was adjusted to the current situation (X-, Y-pixel difference between the corresponding scenes for the tallest building in the test area) else the settings remained the same. For LPS software this time subareas for three object types were identified. This allows for specific search window sizes depending on the land cover in the study area. This was done to improve the DSM results which were not satisfying for the Nairobi test case. Two tests were conducted:

- Quantitative comparison with the reference DSM
- Qualitative visual comparison

For the quantitative accuracy assessment a high resolution reference DSM was available with a vertical accuracy of 0.25m for 89.8% of the area. 5.5% of the area has an accuracy between 0.25 and 1m. The reference DSM was resampled to a horizontal resolution of 0.5m; the original DSM has a horizontal accuracy of 0.2m.

7.2.1 Quantitative Accuracy Assessment

Due to different orthorectification methodologies and models used in the different software packages not only vertical but also horizontal accuracies may vary. This results in horizontal shifts between the DSMs and the reference. The shifts for the test area Graz vary between 3.06m and 3.27m.

For the vertical accuracy comparison the root mean square error (RMSE) and mean absolute error (MAE) were calculated. First, an overall comparison was done by comparing all areas with the reference. Second, the vegetated areas were masked out of the generated DSMs to compare them correctly with the reference DSM.

Vertical Accuracy Assessment [m]	PCI	ENVI	RSG	LPS
RMSE (All areas)	3.91	12.97	3.56	3.05
RMSE (Excluding vegetation)	3.05	14.01	3.25	2.96
MAE (All areas)	2.62	6.85	2.63	2.24
MAE (Excluding vegetation)	2.20	7.69	2.55	2.28

Table 12: Vertical accuracy assessment results comparing the reference DSM with the generated DSMs.

The results (Table 12) show that the DSM calculated with LPS achieved the highest accuracy compared to the reference DSM with an RMSE of 2.96m followed by the DSM generated with PCI (RMSE of 3.05m) and RSG (RMSE of 3.25m). The DSM generated by ENVI resulted in a much lower accuracy with an RMSE of 14.01m. When we look at a different accuracy measure, the MAE the results look slightly different. PCI achieved a slightly more accurate DSM with an MAE of 2.20m, followed by LPS with 2.28m and RSG. This indicates that the quantitative error differences are very small between PCI and LPS. Figure 16 - Figure 19 show where the produced DSMs generated most errors compared to the reference DSM.

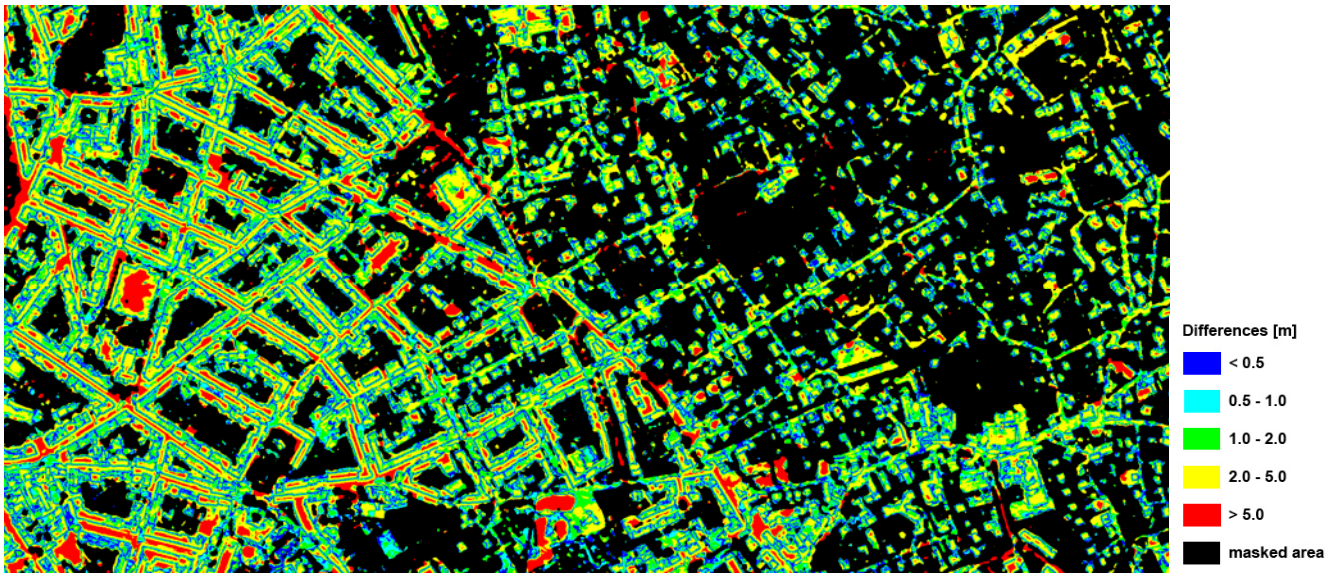


Figure 16: Difference between the reference DSM and the generated DSM with PCI.

In the difference image (Figure 16) the DSM generated with PCI software resulted in the largest errors (in red) estimating the height of large, tall buildings that have a rather complicated structure. Examples are the “Herz Jesu” church and the “Technical University” of Graz. Both building heights were underestimated. Furthermore the ground elevations of roads in the western city centre are erroneous. Errors larger than 5m were achieved for roads between tall buildings in the city centre. The best height accuracies (blue, turquoise) were achieved for small residential houses and some of the multi-storey buildings in the city centre as well as open impervious areas.

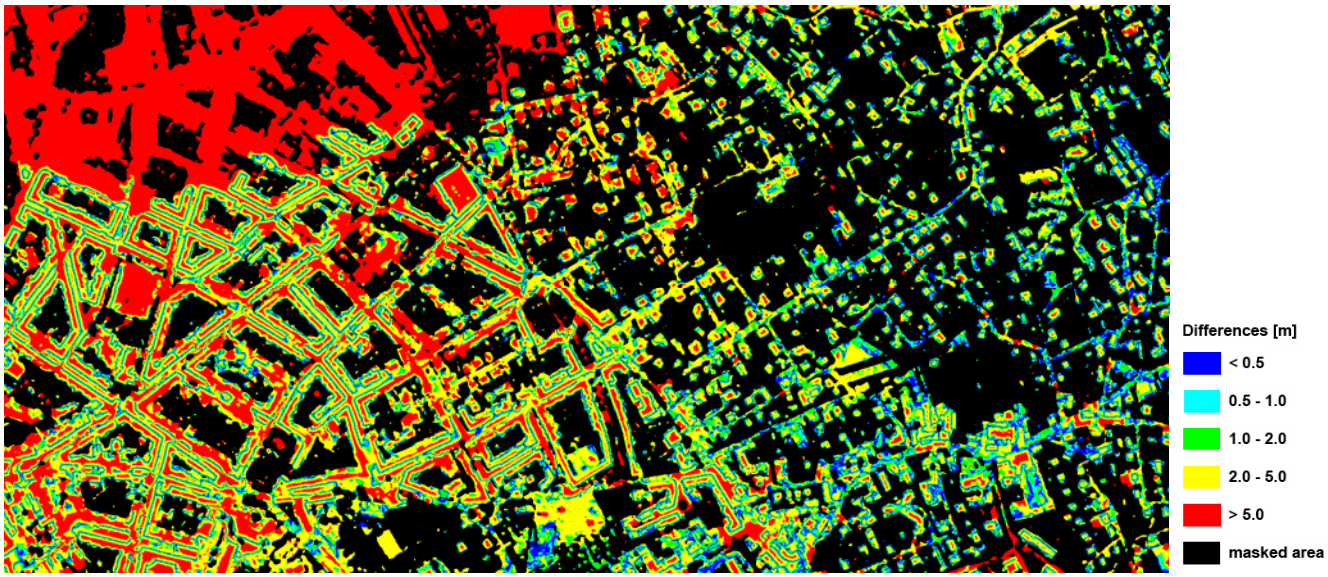


Figure 17: Difference between the reference DSM and the generated DSM with ENVI.

In Figure 17 the difference image of ENVI is displayed. The red area in the north indicates the erroneous cloud created by ENVI. Most multi-storey buildings and roads in between show large errors.

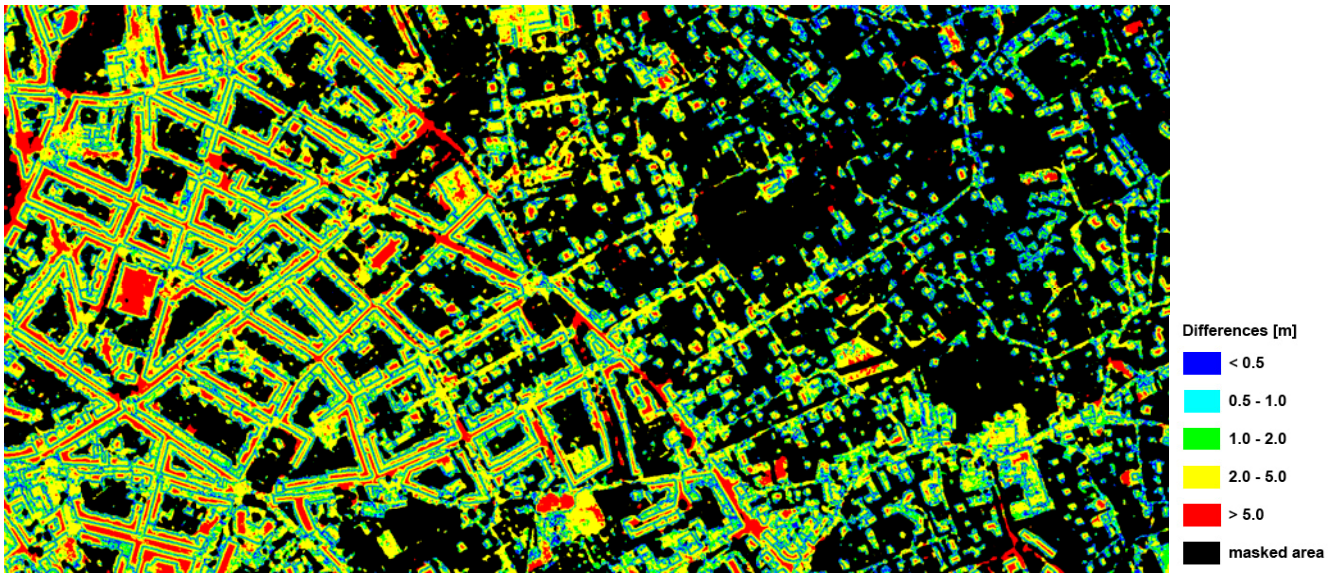


Figure 18: Difference between the reference DSM and the generated DSM with RSG.

The difference image of RSG shows that the Eastern residential buildings were estimated well. Also RSG, similar to PCI, had problems generating the correct building heights for the “Technische Universität” and the “Herz Jesu” church. Both difference images are very similar although it seems that the difference image created with PCI contains more blue areas than the difference image created with RSG which is dominated more by green areas which would also confirm the quantitative results.

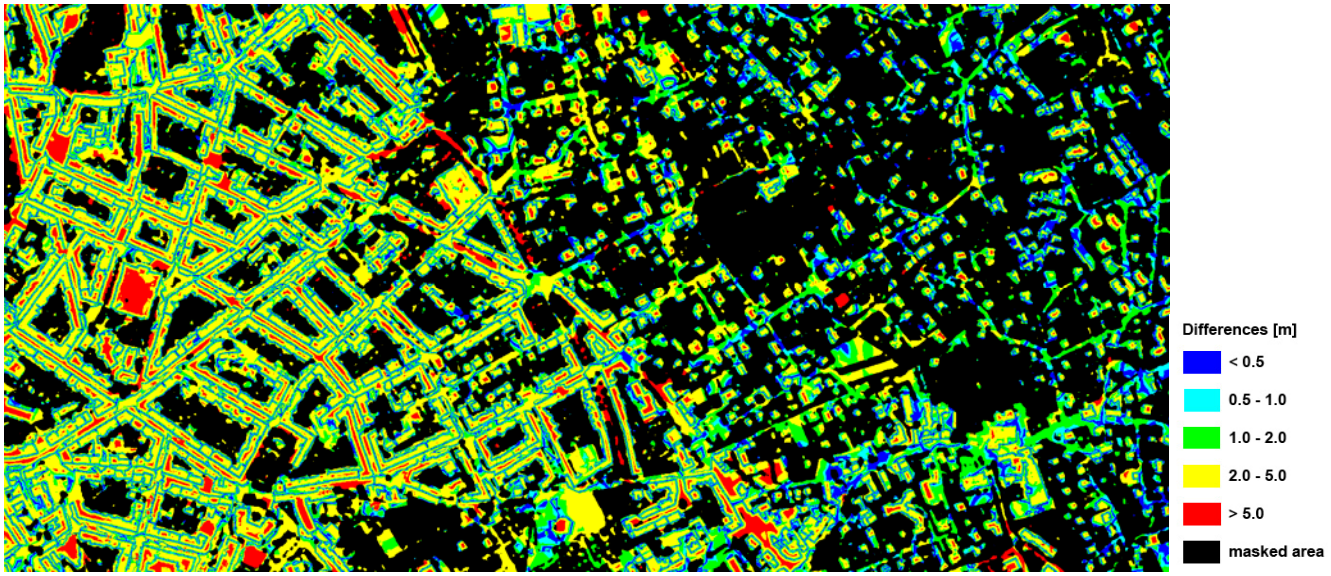


Figure 19: Difference between the reference DSM and the generated DSM with LPS.

The difference image created with LPS shows good results for the residential buildings in the East of the test area. LPS was the only software which recognised the “Herz Jesu” church. However, it failed extracting the heights of the “Technische Universität”. Also the difference image created with LPS is very similar to the difference image of RSG and PCI. The main weakness of LPS software, producing fuzzy object outlines and large interpolation areas is not apparent in this quantitative analysis. Assuming that the mean building heights are extracted by using an available cadastral building outline vector layer the calculated building heights would result in accurate estimates.

This quantitative analysis showed that if DSMs are generated from stereo IKONOS data approximately an error representing “one floor” has to be considered. The software packages PCI, LPS and RSG achieve results with very similar quantitative accuracies. In the next chapter the performance of the software packages in detail extraction will be evaluated.

7.2.2 Visual Evaluation

The four generated DSMs and the reference DSM are shown below in Figure 20 to Figure 24. On the left of the reference DSM, representing the city centre, the multi-storey buildings with inside courts are visible and towards the right, further away from the city centre, residential houses and residential apartment buildings are present. The ground elevation rises in the test area from Southwest to Northeast.

The DSM generated with PCI shows that the ground elevation was estimated well. The multi-storey buildings were extracted well and the outlines are clearly visible. However, some multi-storey row houses weren't clearly extracted and roads in between are not visible in the generated DSM. The residential areas towards the East were extracted well. Some objects in flat non built-up areas can be explained by trees.

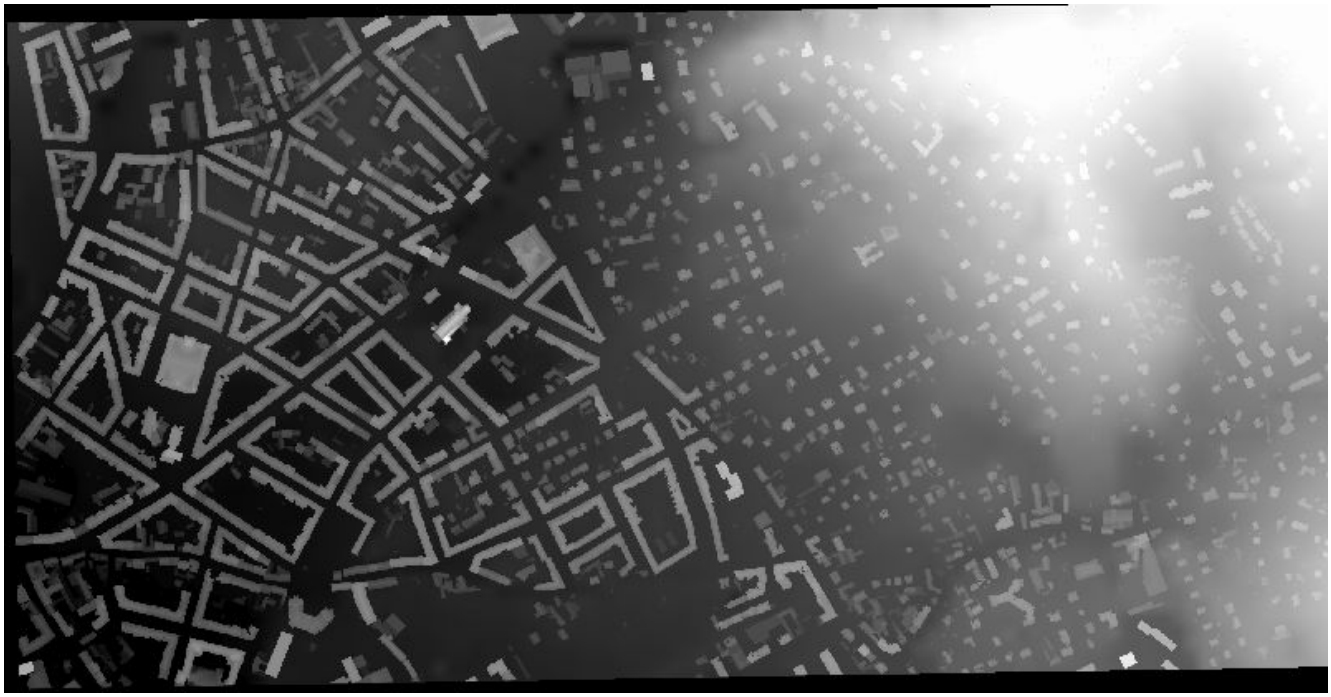


Figure 20: Reference DSM with a horizontal resolution of 0.5m.

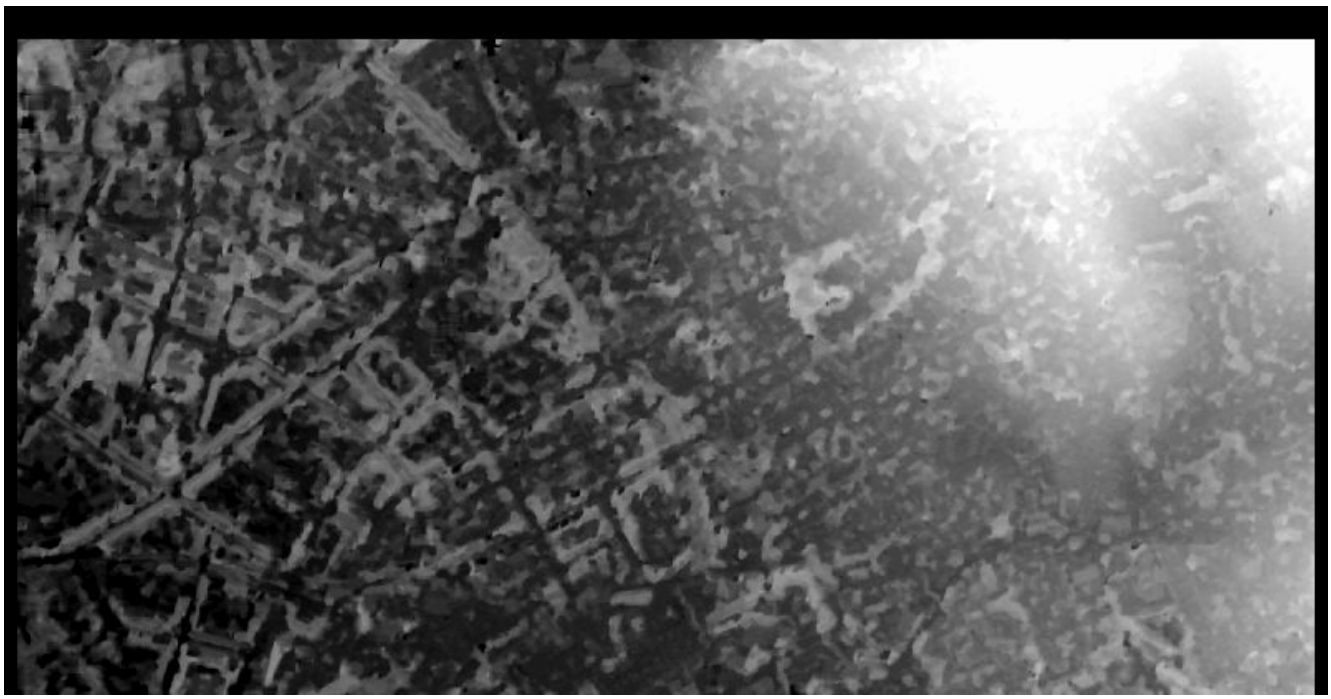


Figure 21: DSM generated with PCI Geomatica software.

Figure 22: DSM generated with ENVI software.

The DSM created by ENVI (Figure 22) generated an erroneous “mountain” in the North of the city centre. In the Southwest the multi-storey buildings were extracted as well as the residential houses in the East of the test area. However, the buildings were not extracted clearly compared to the DSMs extracted by the other software packages.

The DSM created by RSG (Figure 23) is representing well the rising ground elevation towards the East of the test area. The multi-storey buildings were detected but not as clearly as in the DSM generated by PCI. The residential houses were extracted well. In general all building outlines seem to be fuzzy. Also in this DSM some multi-storey buildings are connected with each other and the roads were not clearly detected.

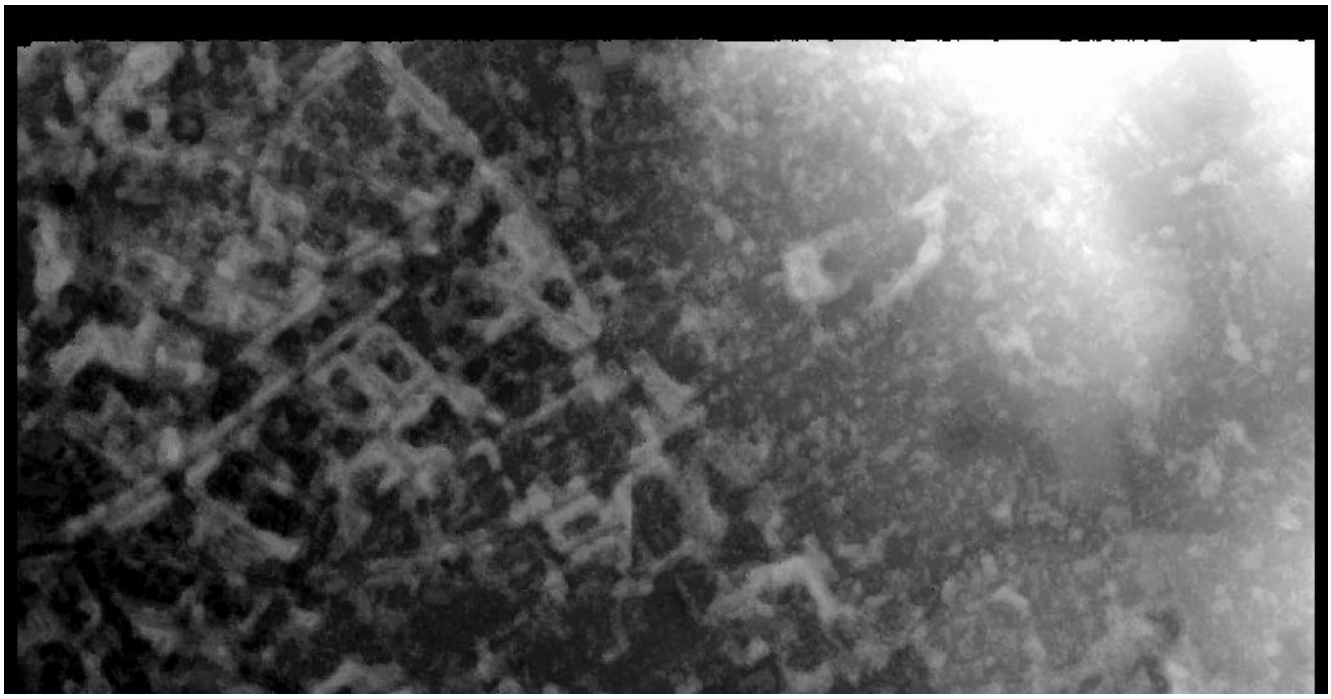


Figure 23: DSM generated with RSG software.

The DSM generated with LPS achieving the best quantitative result with an RMSE of 2.96m produced a DSM with a lot of fuzzy interpolated areas and buildings are not clearly distinct (Figure 24). Moreover the single residential houses are extracted with a low level of detail. Several settings in the software LPS were tested in order to improve the scale of detail of the DSM but without success.

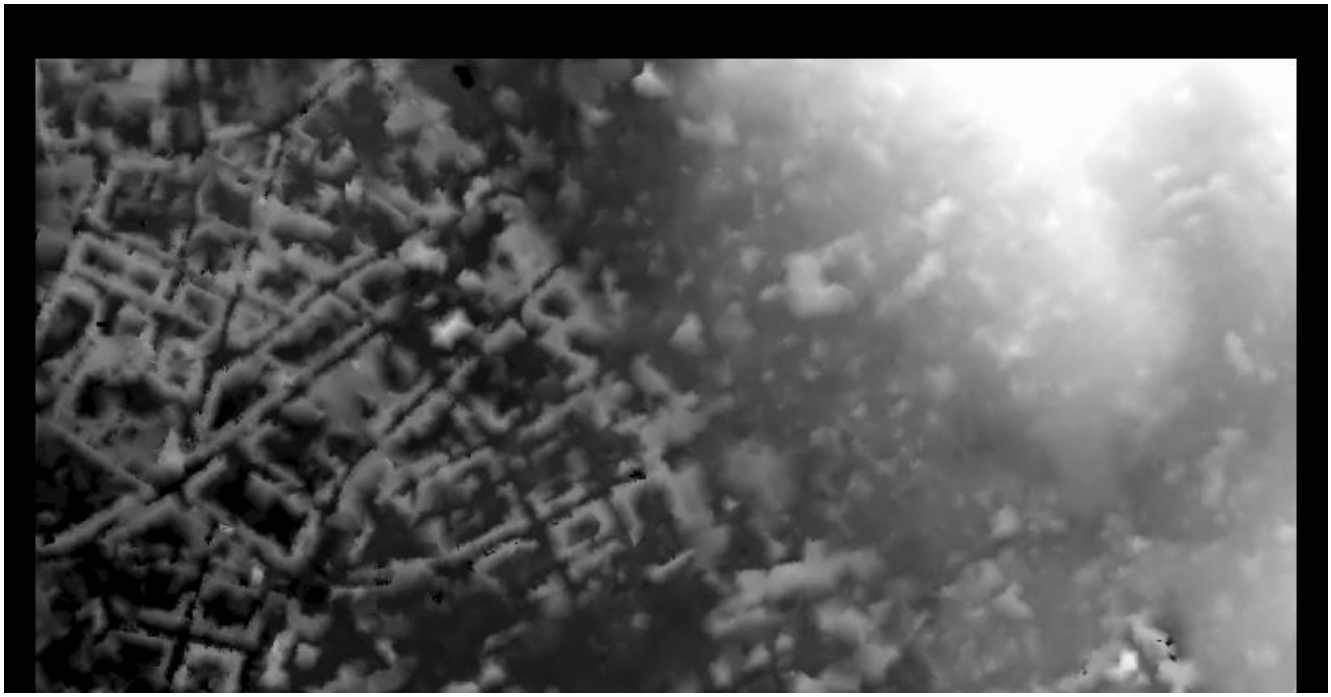


Figure 24: DSM generated with LPS software.

7.2.3 Discussion

The quantitative accuracy assessment for the test case of Graz has shown that the best vertical estimation results were achieved with the software packages of LPS and PCI followed by RSG. The vertical MAE for built-up and impervious areas was 2.20m for PCI, 2.28m for LPS and 2.55m for RSG respectively. The RMSE was 3.05m, 2.96m and 3.25m respectively. Besides the vertical also the horizontal error should be considered depending on the different orthorectification methodologies applied (rigorous physical or RF based models). The shift compared with the reference DSM was between 3.06m and 3.27m.

The qualitative, visual DSM evaluation has not confirmed the quantitative results. LPS with the best quantitative accuracy created fuzzy building outlines and contains low details in areas with smaller objects. PCI and RSG both produced DSMs with clear building outlines. They both are able to extract high details in areas with small buildings. Both DSMs create some errors by connecting multi-storey buildings rooftop heights ignoring the roads ground elevation laying in between. For the residential area in the East it seems that the structures are most clearly extracted by PCI but RSG is detecting better the finer details. Besides achieving the largest error in the quantitative analysis due to an erroneous mountain in the North of the test area ENVI also had problems in extracting correct multi-storey buildings outlines in the denser city area. It achieved good visual results with high details for rather small buildings.

Summarizing it can be said that the recommendations made for the Nairobi test area were confirmed by the quantitative and qualitative accuracy assessment done in the Graz test area. Both, PCI and RSG performed well or achieved at least acceptable results in both the quantitative and qualitative analysis. They both are recommended for digital surface model generation over built-up areas and settlements. Although LPS achieved the best quantitative accuracy it failed in creating DSMs with high details and clear building outlines. This is essential for building height extraction if the building outlines have to be still extracted from the data itself and are not available from cadastral offices as vector data layer. At last, ENVI showed a weak performance in creating large erroneous elevations in the test areas. It failed in achieving good quantitative results for both buildings and ground elevations. However, it should be mentioned that it successfully extracted fine structures of very small buildings in the Nairobi test area.

8 Building Height Extraction – Methodology Development

In order to know the height of an object the difference between a DSM and a DEM has to be known. This height difference can be extracted in various ways such as applying sophisticated interpolation or filtering techniques. The different approaches can be grouped into direct and indirect derivation techniques.

The indirect method is based on the resulting difference between a DSM and a DTM. In this case first a DTM has to be created from a DSM which is generally done by interpolating known altitude ground points around objects. The challenge is to choose correct ground points and then to interpolate ground height in the whole scene. As final surface depends on the selected ground points, their selection is crucial as well as is their amount and distribution. Large "no data" areas can lead to interpolation artefacts and erroneous elevation heights.

The direct method is based on sophisticated filtering methodologies such as morphological operations where the structures representing object heights are directly derived from the DSM and no interpolation is required.

After the derivation of the difference height between DSM and DEM the building outlines have to be known to be able to calculate a mean building height. Building vector layers are available for most first world countries and for some cities in the developing world. However, if this kind of information is not available it has to be either manually digitized or automatically extracted. We are trying to develop such an automatic building outline extraction methodology by using a watershed segmentation approach. An overview of the possible methods and their workflow processes are illustrated in Figure 25. Both methodologies were applied in the Nairobi test case as well as in the Graz test case. The theoretical aspects as well as first results are presented.

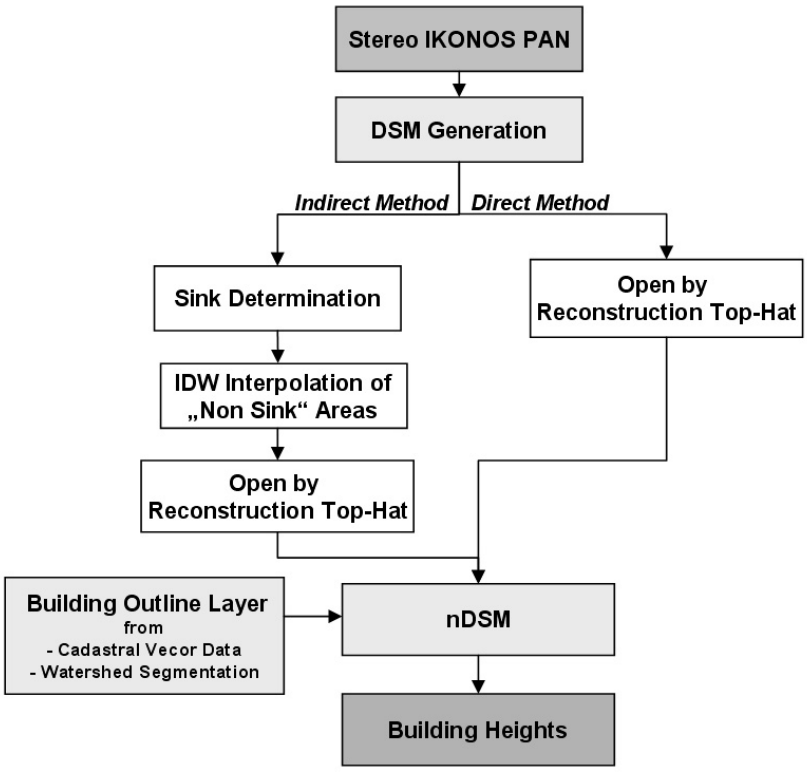


Figure 25: Overview of the possible methods and their workflow processes.

8.1 Direct Object Height Calculation – Based on Morphological Analysis

The core is the object height extraction by a morphological operation. The fundamental operators in mathematical morphology are *erosion* and *dilation* (Soille, 2003). When mathematical morphology is used in image processing, these operators are applied to an image with a set of a known shape, called a structuring element (SE). The application of the erosion operator ε_N to an image gives an output image f(p), which shows where the structuring element N fits the objects in the image $SE \equiv N_G(p)$. It can be written as follows:

$$\varepsilon_N f(p) = \{\wedge f(p') \mid p' \in N_G(p) \cup f(p)\} \quad (1)$$

where the erosion ε_N of the grey level function using the structuring element *N*, is defined by the infimum of the values of the grey level function in the neighbourhood.

On the other hand, the application of the dilation operator δ_N to an image gives an output image *f(p)*, which shows where SE hits the objects in the image. It is defined by the supremum of the neighbouring values and the value of *f(p)* and can be written as:

$$\delta_N f(p) = \{\vee f(p') \mid p' \in N_G(p) \cup f(p)\} \quad (2)$$

The erosion and dilation operators are dual but noninvertible, in general. All other morphological operators can be expressed in terms of erosion and dilation.

Two commonly used morphological operators are *opening* and *closing*. The idea behind opening γ is to dilate an eroded image in order to recover as much as possible of the eroded image.

$$\gamma_N f(p) = \delta_N \varepsilon_N f(p) \quad (3)$$

In contrast, the idea behind closing φ is to erode a dilated image in order to recover the initial shape of image structures that have been dilated.

$$\varphi_N f(p) = \varepsilon_N \delta_N f(p) \quad (4)$$

The filtering properties of the opening and closing operators are based on the fact that not all structures from the original image will be recovered when these operators are applied. It is a common practice to use the opening and closing transform in order to isolate bright (opening) and dark (closing) structures in images, where bright/dark means brighter/darker than the surrounding features in the images (Benediktsson et al., 2003). If a greyscale image is interpreted as a digital surface or topography, then the opening cuts the objects or peaks. In contrast, closing fills holes and valleys that are smaller than the SE. In order to isolate features with a thinner support than a given SE, a widely used technique is to take the residuals of the opening, closing, and original images, by a morphological transformation called *top-hat* $\Gamma f(p)$ and *inverse top-ha t* $\Gamma' f(p)$, where:

$$\Gamma f(p) = f(p) - \gamma_N f(p) = f(p) - \delta_N \varepsilon_N f(p) \quad (5)$$

and

$$\Gamma' f(p) = \varphi_N f(p) - f(p) = \varepsilon_N \delta_N f(p) - f(p) \quad (6)$$

(Serra, 1982).
Here, the chosen approach for the opening and closing calculation uses a non-Euclidean metric known as *filtering by reconstruction* (Crespo et al., 1995). The reason for using the reconstruction approach is that it has proven to have a better shape preservation than classical morphological filters. In fact, reconstruction filters introduce nominally no shape noise, since the shape of the structuring element used in the filtering are adaptive with respect to the structures present in the image itself (Soille and Pesaresi, 2002). The greyscale reconstruction $\rho^I(J)$ of an image *I* (mask) from *J* (marker) is obtained by the iterative use of an elementary greyscale geodesic dilation of *J* inside *I* until idempotence is achieved

$$\rho^I(J) = \bigvee_{n \geq 1} \delta^I_{(n)}(J) \mid \delta^I_{(n)} = \delta^I_{(n+1)} . \quad (7)$$

In this study, the structures representing the object heights are derived by applying an *open by reconstruction top-hat* operator to the DSM. As SE a disk with a radius of 50m is used, after testing different disk sizes with a radius of 25m, 50m and 75m. The disk size should be approximately as large as the buildings present in the area. The resulting object heights layer is illustrated in Figure 26.

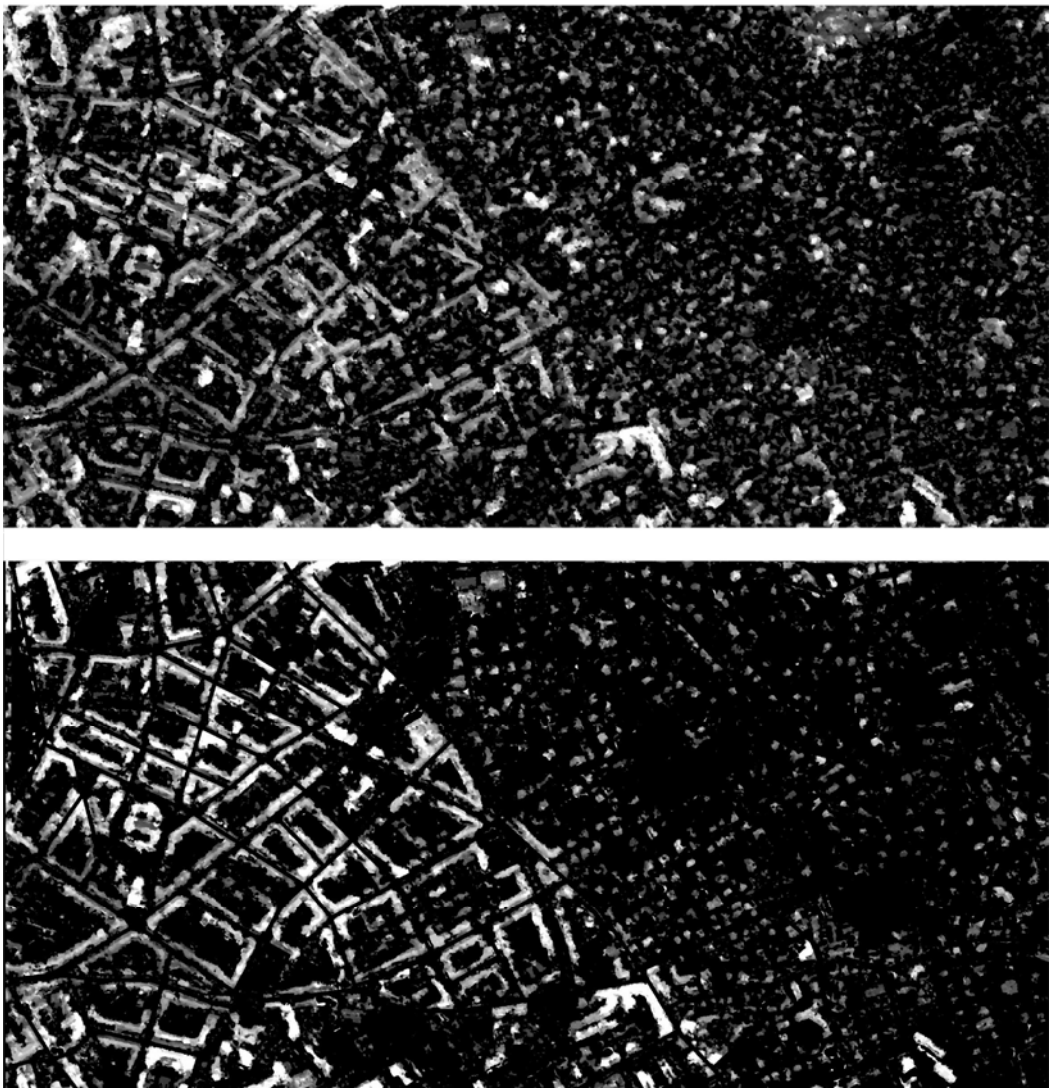


Figure 26: Object height layer (above), building height layer, NDVI and road masked (below).

8.2 Indirect Object Height Calculation – Based on Filtering and Interpolation

The indirect method is based on the creation of a DEM from the generated DSM by finding ground elevation, non built-up or forested areas and interpolating the built-up or forested areas with an inverse different weighting (IDW) interpolation algorithm. The ground elevation areas can be found by defining all sinks in the DSM (Figure 27) with the assumption that they are actually representing a correct ground elevation. The remaining areas are interpolated then by using the sinks' elevation values (Figure 28).

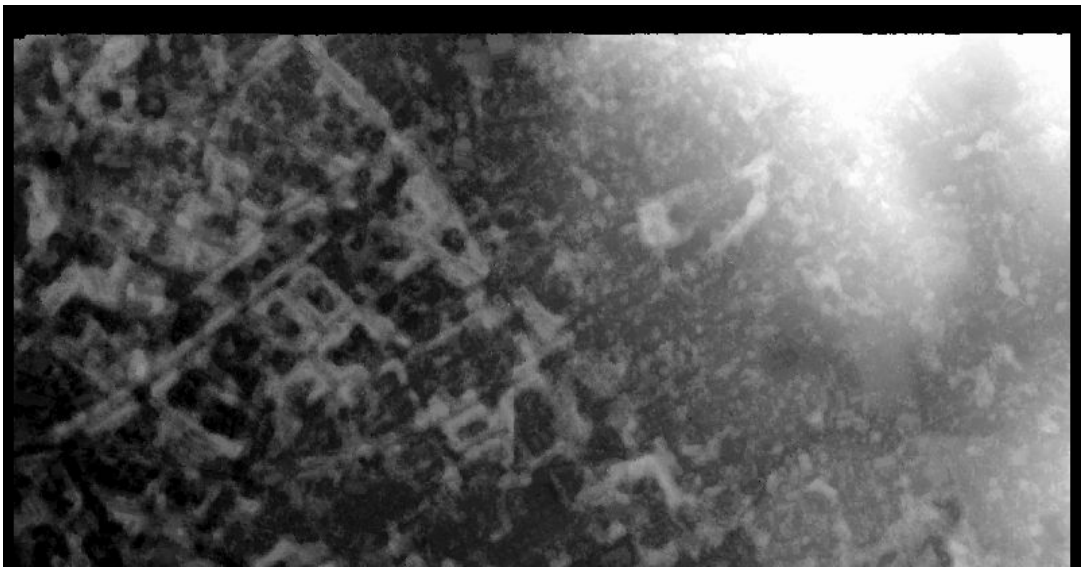


Figure 27: Original DSM generated with RSG software.

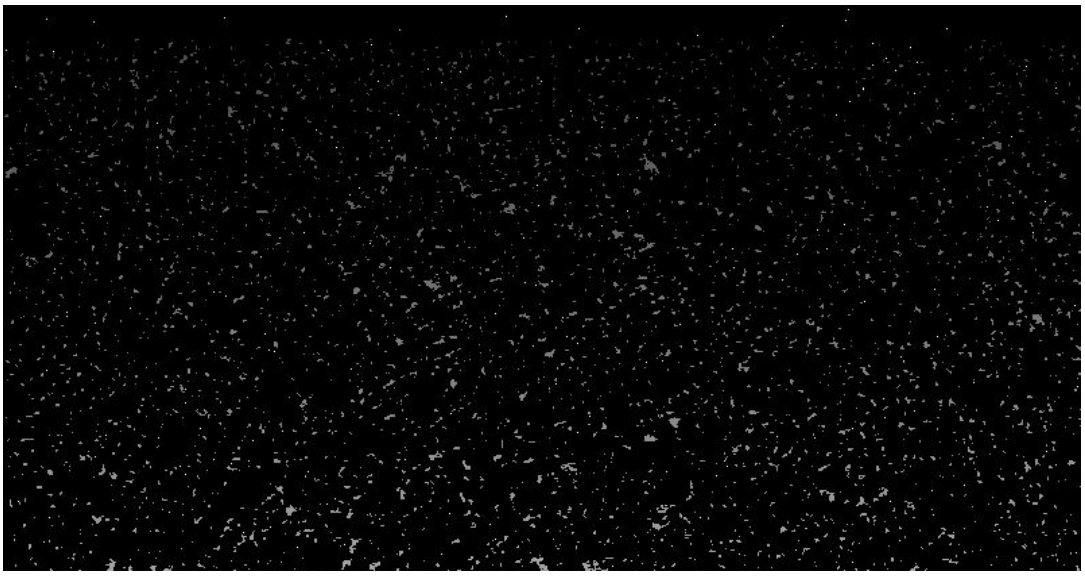


Figure 28: Sinks identified in the original DSM. The black areas are no data pixels which are interpolated.

Several interpolation algorithms were tested such as kriging, spline, natural neighbor and IDW. The IDW algorithm obtained the best results (Figure 29). It determines cell values using a linearly weighted combination of a set of sample points. The weight is a function of inverse distance. The surface being interpolated should be that of locationally dependent variable. At last, the created DEM can be filtered, if necessary, by applying a morphological erosion to eliminate erroneous elevation peaks and sinks (Figure 30). Finally the DEM is subtracted by the DSM and the difference remains as object height layer also called nDSM (Figure 31).

The building heights can now be extracted by overlaying a building outline vector layer and calculating a mean height for every building. If a cadastral building outline vector layer is not available it can be extracted from the satellite data and DSM itself by watershed segmentation. Research developing an appropriate methodology is still ongoing. The methodology and first experimental results are presented next.

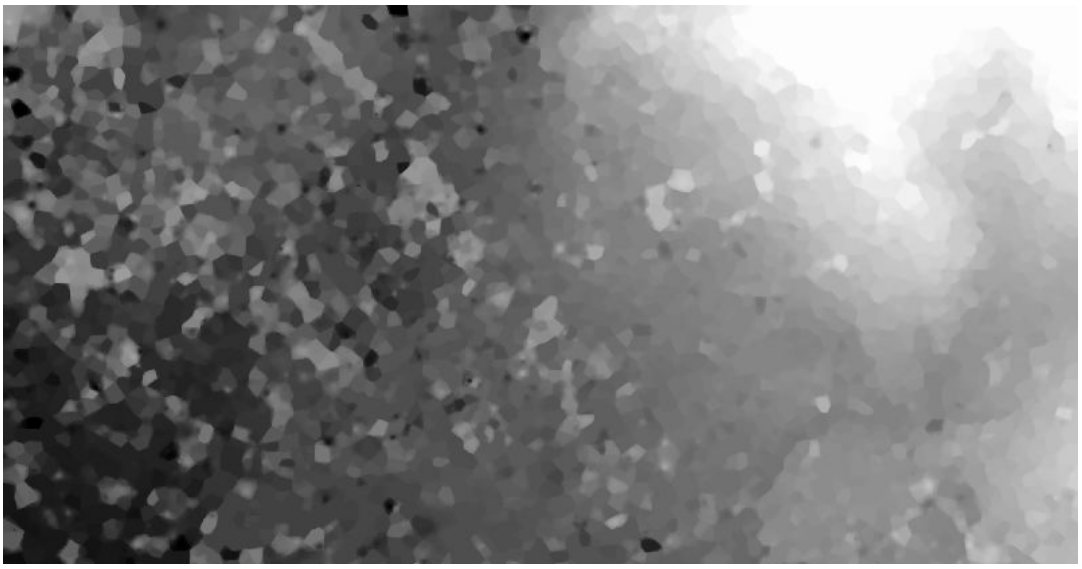


Figure 29: DEM created by IDW interpolation.

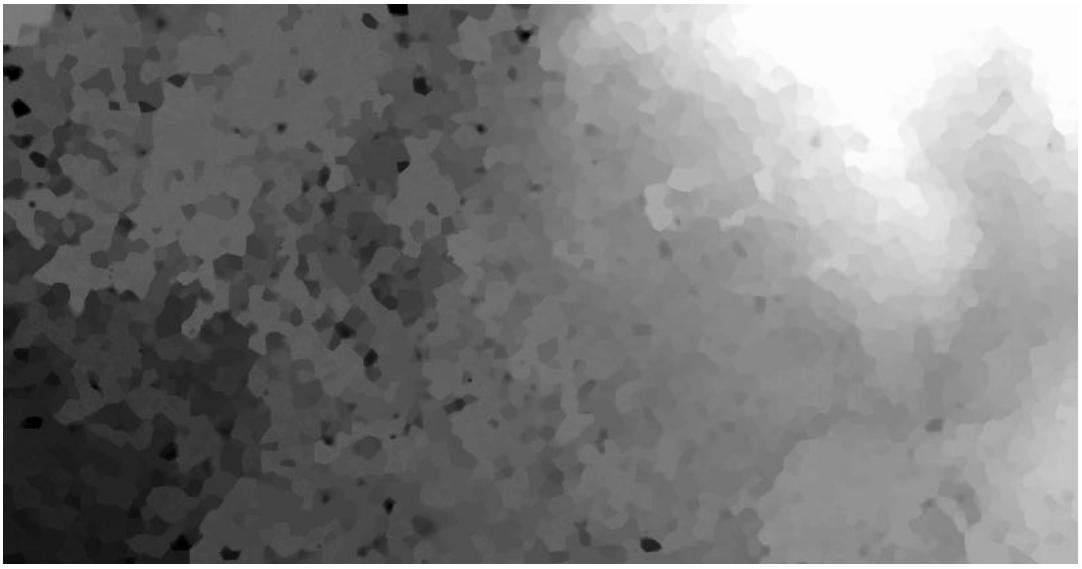


Figure 30: Filtered DEM.

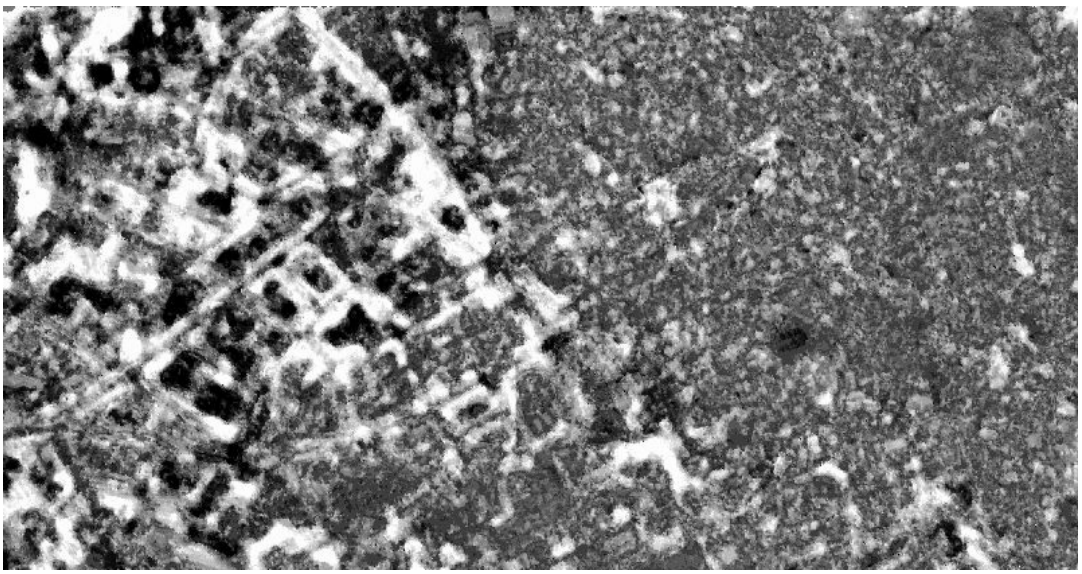


Figure 31: Difference between DSM and DEM resulting in an object height layer.

8.3 Watershed Segmentation

Watershed segmentation combines region growing and edge detection techniques by grouping the image pixels around the regional minima of an image and the boundaries of adjacent groupings are precisely located along the crest lines of the gradient image. It appears to be a very powerful segmentation tool. Provided that the input image has been transformed so as to output an image whose minima mark relevant image objects and whose crest lines correspond to image object boundaries the watershed transformation will partition the image into meaningful regions or objects respectively (P. Soille, 2003). The standard approach in watershed segmentation causes severe over-segmentation due to the presence of irrelevant local minima or maxima. To overcome this problem a marker is introduced (Meyer and Beucher, 1990). The basic idea behind the marker-controlled segmentation is to transform the input image in such a way that the watersheds of the transformed image correspond to meaningful object boundaries. Generally this is done by morphological filtering (in this case by opening and closing) of the image. Then the filtered gradient is thresholded. This approach assumes that the region of interest for detection is large and homogeneous, which is mostly not the case for DSMs and VHR satellite imagery. Thus the watershed segmentation still results in over-segmentation and erroneous segments which can influence the shape and height accuracy of the extracted building heights.

In the applied watershed segmentation an opening and closing by reconstruction is performed using a disk of size of one pixel for the building height layer and a disk of size seven pixels for the PAN imagery as the structuring element. The gradient area for the closing is one pixel for the building heights layer and six pixels for the PAN imagery. The threshold value is set to zero for both image segmentations.

The idea is to intersect and update the watershed segmentation of the PAN imagery containing the edge-sharpened boundaries of the buildings with the watershed segmentation of the DSM itself containing the heights of the buildings. Therewith building outline information contained in both layers, the PAN imagery and the building height layer are taken into account.

At last, the mean building heights are calculated assuming that every segment is representing parts of a building or ideally a single building. This research is ongoing and first results are illustrated in Figure 32 where the extracted segments and the reference building outlines are illustrated.

Figure 32: Segmentation layer (grey) overlaid with the reference building outlines (black). Left: city multi-storey buildings, right: single residential buildings.

Figure 32 shows that most buildings are detected as such but there are problems to extract the multi-storey buildings as one building segment. The single residential buildings are detected well and segmented partially. The segments in the lower west area of right image are houses not contained in the reference building layer therefore no reference shapes are available. The segments in the lower east corner are erroneous. They were not eliminated during the vegetation masking step and are as building segmented vegetation areas. Overall the results are promising. Most buildings are detected. Future work will focus on the improvement of the building shapes.

8.4 Experimental Quantitative Results

8.4.1 Reference Building Heights vs. DSM Building Heights – Difference Analysis

In this analysis the building heights were extracted by using the building outline vector layer derived from the reference DSM. The extracted mean heights of the reference were then compared with the extracted mean heights from the DSM generated with RSG software by calculating a difference image.

The mean difference or error for all polygons was statistically analyzed. In Table 13 the results for all 1247 polygons are summarized. Of the 1247 polygons for 433 polygons the difference or error was smaller than two meter and for 652 polygons smaller than three meters which equals to one floor. 40 polygons are afflicted with an error larger than 10 meters.

Number of polygons: 1247	Min	Max	Mean	Stdev
Reference - DSM	-20.38	24.74	0.78	4.69

Table 13: Mean height difference between the reference and the generated DSM by RSG based on the reference building outline layer.

A visual inspection shows that large errors were calculated for very small buildings. An explanation is that they are partially covered by tree crowns or are very high multi-storey buildings and therefore their height was not correctly extracted. The same applies to the “Herz Jesu” church, which was only detected by the DSM generated with the LPS software package. Figure 33 illustrates the calculated height differences for the different building polygons.

Figure 33: Mean building height differences between the generated DSM by RSG and the reference.

8.4.2 Building Height Extraction – Comparison of direct and indirect method

Next the two building height extraction methodologies were tested by comparing them with the reference building heights. The calculated statistics show that the indirect interpolation-based methodology achieves more accurate building heights than the direct morphological-based methodology. The MAE is 4.53m for the indirect and 5.97m for the direct building height methodology.

Difference Analysis [m]	Min	Max	Mean	Stdev	RMSE	MAE
Reference – Indirect Method	-18.20	25.78	3.91	4.04	5.63	4.53
Reference – Direct Method	-16.20	27.83	5.74	4.12	7.07	5.97

Table 14: Statistical analysis of the differences between the reference building heights and the building heights derived by applying the indirect and direct method to the DSM generated with RSG software.

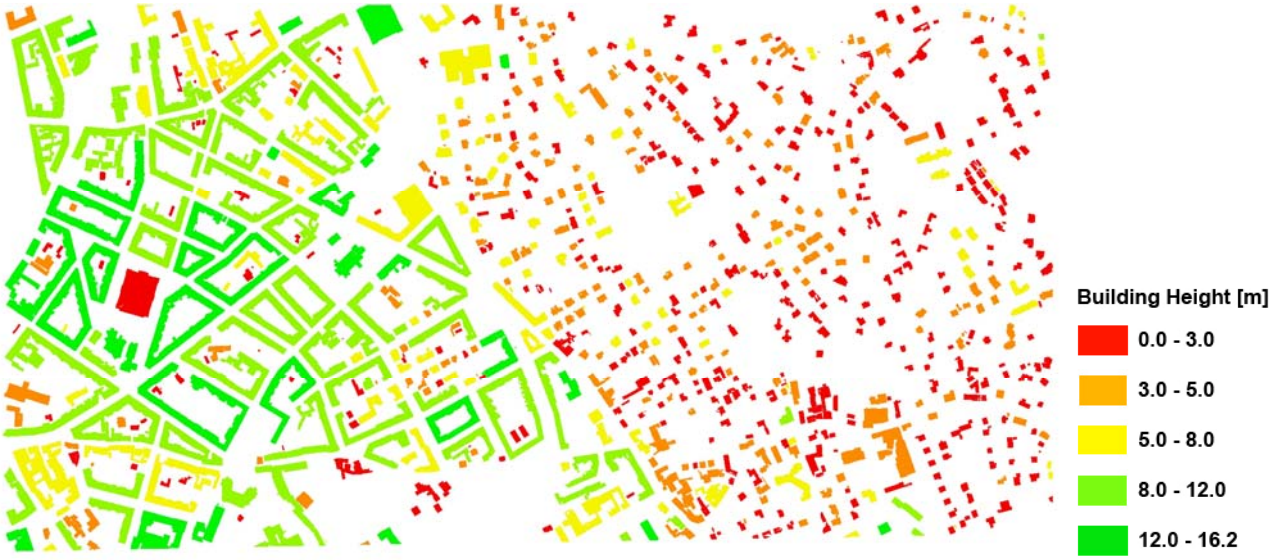


Figure 34: Building height derived from the reference DSM.

Figure 35: Building height derived from the direct morphology-based method applied on the DSM generated with RSG.

Figure 36: Building height derived from the indirect interpolation-based method applied on the DSM generated with RSG.

Figure 35 and Figure 36 represent the building height maps calculated based on the two building height extraction methodologies and the DSM generated with RSG software. Comparing their building height classes they give an overview of where and for what building types the errors are largest compared with the reference building heights shown in Figure 34. It can be seen that the direct morphological-based approach underestimates all multi-storey buildings in their heights. The discrepancies are smaller for the residential single buildings. The indirect interpolation-based approach underestimates building heights as well but with a much smaller height error. It should also be considered that the used DSM, generated by RSG had a slightly larger MAE than the best DSM generated with PCI. Therefore, if using the most accurate DSM the results may improve slightly.

8.5 3D Visualisation

Knowing the third dimension buildings can be visualized in 3D. Today the demand for the generation and realistic visualization of 3D urban environments is growing in many application fields. Representing the environment as real as possible (e.g. virtual fly through or print-outs) can be a valuable information for efficient disaster simulations, disaster situation analysis, vulnerability and risk analysis not only for scientists but also for decision makers. Examples of Nairobi showing the RGB composite of QuickBird draped over the building heights extracted with the indirect method can be seen in Figure 37 and Figure 38. The building outlines were extracted applying the watershed segmentation to the building height layer and the IKONOS PAN data and intersecting the resulting segment layers.

Figure 37: 3D visualization of a residential area of Nairobi.

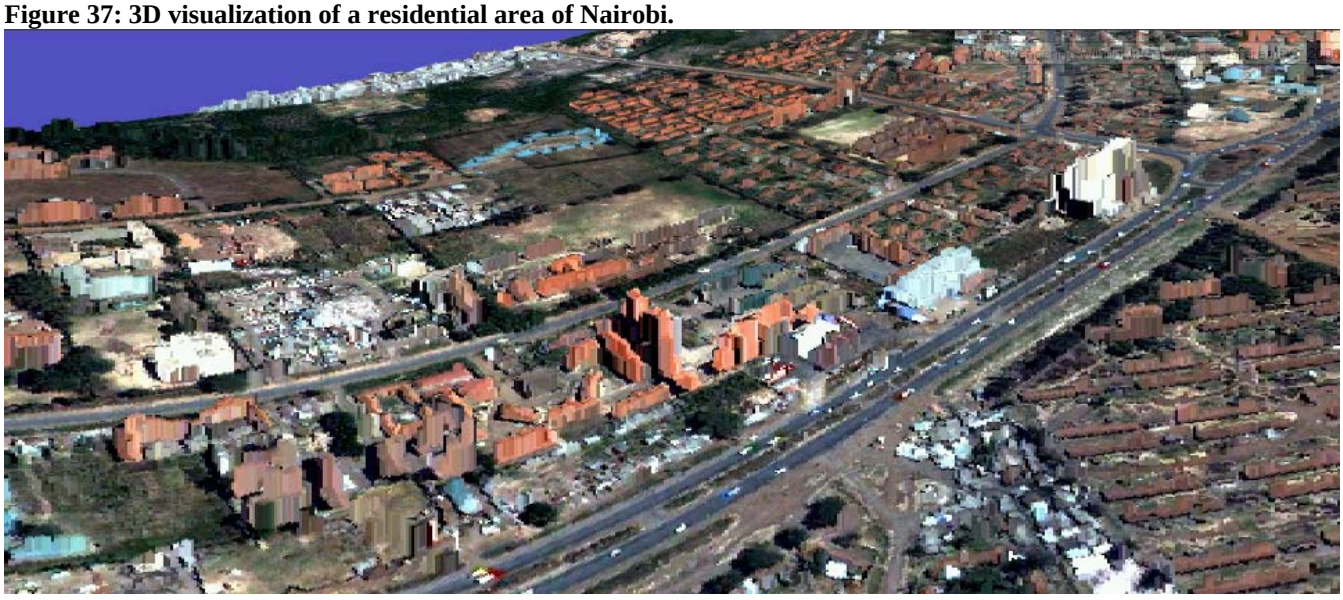


Figure 38: 3D visualization of Nairobi.

Figure 39 and Figure 40 are examples of the city of Sana'a. The mean building heights were calculated for every building by applying the indirect building height extraction methodology. The layer was combined with a digital elevation model and visualized in 3D. The building outline information was available as digital vector layer which we had received from the city of Sana'a. The examples illustrate well how dense the buildings are standing next to each other in the old town of Sana'a, looking in northeast direction (Figure 40).

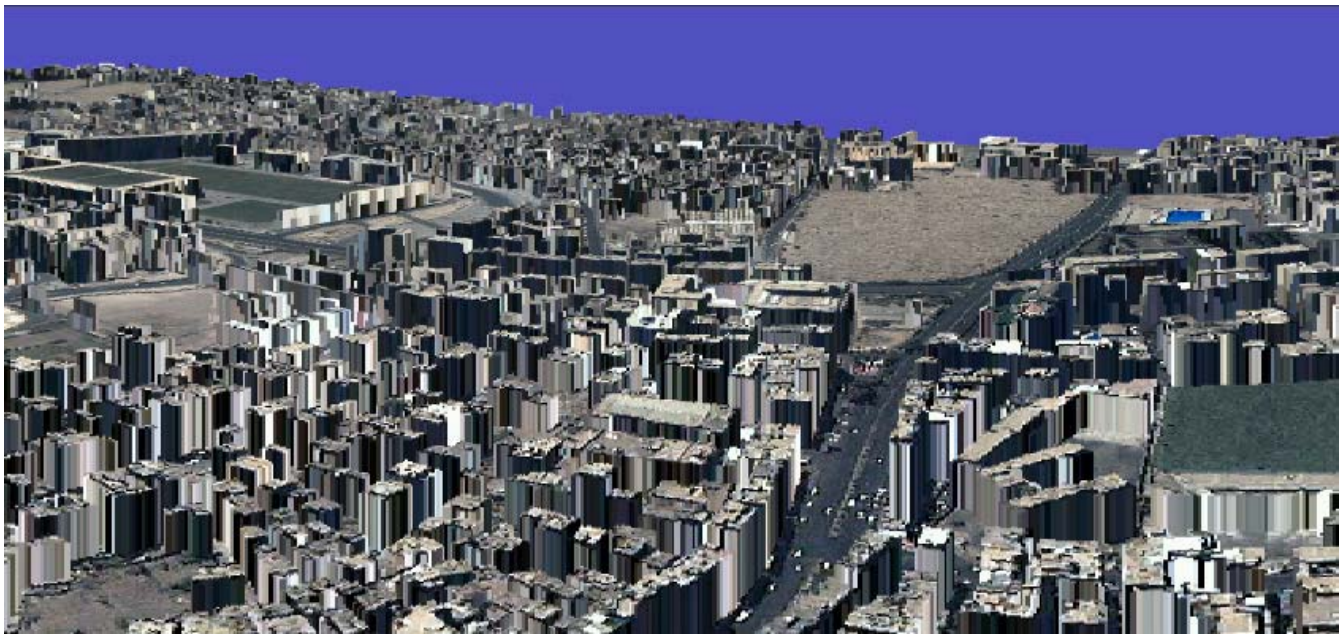


Figure 39: 3D visualization of Sana'a, looking towards the South of the city.

Figure 40: 3D visualization of Sana'a, looking towards the North of the city. In the Northeast the old town of Sanaa is clearly visible with its small and dense buildings.

8.6 Discussion

To extract building heights from stereo satellite data two problems have to be solved. First, the object height information has to be derived from the generated DSM. Two methodologies were presented to derive the object height layer: an indirect and a direct methodology. Second, the building outlines have to be delineated and extracted. A possible approach was proposed based on watershed segmentation.

The results received with the two tested methodologies are promising. A mean absolute error of 4.53m and 5.97m respectively was achieved. Buildings with a maximum height of 8m in the peri-urban area of Graz were estimated with mean height errors of few meters representing one building floor or less if a building floor is assumed to be of approximately 3m. However, tall buildings with sizes of more than 8m in the city centre of Graz, where buildings are standing close to each other, were underestimated by several meters representing two or more building floors. These discrepancies have to be further analyzed. In case they are constant or linear the addition of an offset could be integrated into the current extraction methodologies.

A building outline extraction approach based on watershed segmentation and preliminary results were presented. The methodology successfully detected most buildings. Problems occur where buildings have complicated outlines. The extracted shapes of most building outlines are approximated and not representing the generally rectangular shapes of buildings. The next working steps will focus on the adjustment of the morphological segmentation approach, the integration of spectral information derived from the satellite imagery and on a possible post-processing of the resulting segments.

9 Overall Conclusion

The orthorectification accuracy assessments done for the Nairobi and Graz datasets have shown that the rigorous physical model achieved the best accuracies below (QuickBird) or very close to (IKONOS) pixel size. It proved to be stable for both IKONOS and QuickBird data using both few and a large number of GPCs. It is therefore recommended to use the rigorous physical model for orthorectification of VHR satellite data if high precision GCPs are available and a high geometric accuracy of the data is necessary. The RF model is still a viable alternative when high accuracy GCPs are very limited or not available. If the topography is rather flat in a dataset a 0-order polynomial adjustment with RF model orthorectification might be sufficient but a 1-order polynomial adjustment should be preferred if the terrain is rugged.

Quantitative height accuracy as well as clear building shape extraction is of great importance for the use of DSMs in information extraction for settlement analysis and mapping. Different software packages were therefore tested and the derived DSMs quantitatively and qualitatively analyzed. For both study areas, Nairobi and Graz, similar results were achieved. Overall it can be said that a quantitative vertical accuracy of around 3m RMSE can be achieved when using PCI, RSG or LPS software packages. ENVI created DSMs with larger errors.

The qualitative evaluations have shown that the software packages created DSMs with different levels of detail. The software packages PCI; RSG and ENVI created DSMs with high level of details in urban areas. LPS created fuzzy building outlines and proved not to be suitable for urban DSM generation.

Summarizing it can be said that the recommendations made for the Nairobi test area were confirmed by the quantitative and qualitative accuracy assessment done in the Graz test area. Both, PCI and RSG performed well or achieved at least acceptable results in both the quantitative and qualitative analysis. They both are recommended for digital surface model generation over built-up areas and settlements. However, it has to be considered that computation time for RSG is much higher than for PCI.

To extract building heights from stereo satellite data two problems were addressed: first, the object height extraction from the original DSM, second, the building outline delineation and extraction.

Two methodologies were presented to derive the object height layer: an indirect and a direct methodology. The results are promising. A mean absolute error of 4.53m and 5.97m respectively was achieved. Small and medium-height buildings are estimated with an error of about one floor, taller buildings with a larger error of around two to three floors.

A possible building outline extraction approach based on watershed segmentation was presented. Preliminary results show that the methodology successfully detected most buildings. However, problems still occur where buildings have complicated outlines and the extracted shapes of most building outlines are approximated by circles.

Future work will focus on the improvement of the building height extraction methodologies and on the improvement of the building outline extraction by watershed segmentation.

Acknowledgement

The author would like to thank Joanneum Research, Graz, and especially KH. Gutjahr, for providing the datasets for the Graz test area. The author would also like to thank GMOSS (Global Monitoring for Security and Stability), a network of excellence in the aeronautics and space priority of the Sixth Framework Programme funded by the European Commission's Directorate General Enterprise & Industry.

References

Al-Rousan, N., and Petrie, G., 1998: System calibration, geometric accuracy testing and validation of DEM orthoimage data extracted from SPOT stereopairs using commercially available image processing systems. International Archives of Photogrammetry & Remote Sensing, Vol. 34, 8-15.

Baillard, C., and Dissard, O., 2000: A stereo matching algorithm for urban digital elevation models. Photogrammetric Engineering and Remote Sensing, Vol. 66, 1119-1128.

Baltsavias, E., and Stallmann, D., 1993: SPOT stereo matching for digital terrain model generation. Proceedings of 2nd Swiss Symposium on Pattern Recognition and Computer Vision, Zurich, Switzerland, 1993,61-72

Belgued, Y., Goze, S., Planès J.-P., and Marthon, Ph., 2000, Geometrical Block Adjustment of Multi-Sensor Radar Images, Proceedings of EARSeL Workshop: Fusion of Earth Data, Sophia-Antipolis, France, 26-28 January 2000, 11-16.

Benediktsson, J.A., Pesaresi, M., and Arnason, K., 2003: Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. IEEE Trans. Geosci. Remote Sensing, Vol. 41, 1940-1949.

Caballo-Perucha, M., 2003: Development and analysis of algorithm for the optimization of automatic image correlation. M.S. thesis, Post-graduate University Course Space Science, Karl-Franzens-Universität Graz, Austria.

Canty, M.J., 2007: Image Analysis, classification and Change Detection in Remote Sensing – with Algorithms for ENVI/IDL. CRC, Taylor & Francis, Boca Raton, Florida, USA, 348 p.

Crespo, J., Serra, J., and Schafer, R., 1995: Theoretical aspects of morphological filters by reconstruction. Signal Processing, Vol. 47, 201-225.

Davis, C. H. and Wang, X., 2001: Planimetric Accuracy of IKONOS 1-m Panchromatic Image Products. Proceedings of the ASPRS Annual Conference, St Louis, Missouri, USA, April 23-27.

Dial, G., and Grodecki J., 2002: Block adjustment with rational polynomial camera models. Proceedings of the ACSM-ASPRS Annual Conference/XXII FIG International Congress, Washington D.C., USA, April 19-26.

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. Proc. 20th ISPRS Congress, Istanbul, 2004, 522-528.

Escobal, P.R., 1965: Methods of orbit determination. Malabar, USA: Krieger Publishing Company, 479 p.

Light, D.L., Brown, D., Colvocoresses, A., Doyle, F., Davies, M., Ellasal, A., Junkins, J., Manent, J., McKenney, A., Undrejka, R., and Wood, G., 1980: Satellite Photogrammetry. In: Manual of Photogrammetry 4th Edition, Chapitre XVII, Editor in chief: C.C. Slama, ASP Publishers, Falls Church, USA, 883-977.

Fischer, R., 1999: Untersuchung der Systematik von Vegetationsindicees aus Fernerkundungsdaten am Beispiel Zürich. M.S. thesis, Geographical Institute, University of Zurich, Switzerland, 1999.

Fraser, C. S., Hanley H. B., and Yamakawa, T., 2002: Three-dimensional geopositioning accuracy of IKONOS imagery. Photogrammetric Record, Vol. 17, 465-479.

Grodecki, J., 2001: IKONOS Stereo Feature Extraction – RPC Approach. Proceedings of the ASPRS Annual Conference, St Louis, Missouri, USA, April 23-27.

Hargreaves, D., and Robertson, B., 2001: Review of QuickBird-1/2 and Orbview-3/4 Products from MacDonald Dettwiler Processing Systems. Proceedings of the ASPRS Annual Conference, St Louis, Missouri, USA, April 23-27.

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, Ottawa, Canada July 8-12, Vol. 34 (B4), 637-643.

Kornus, W., Lehner, M., and Schroeder M., 2000: Geometric in-flight calibration by block adjustment using MOMS-2P 3-line-imagery of three intersecting stereo-strips. Bulletin de la Société Française de Photogrammétrie et de Télédétection, Vol. 159, 42-54.

Kristóf, D., Csató, É., and Ritter, D., 2002: Application of High-resolution Satellite Images in Forestry and Habitat Mapping – Evaluation of IKONOS Images Through a Hungarian Case Study. International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences, Ottawa, Canada July 8-12, Vol. 34 (B4), 602-607.

Leberl, F., Maurice, K., Thomas, J.K., and Millot, M., 1994: Automated Radar Image Matching Experiment. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 49, No. 3, 19-33.

Lee, J.-B., Huh, Y., Seo, B., and Kim, Y., 2002: Improvement the Positional Accuracy of the 3D Terrain Data Extracted From IKONOS-2 Satellite Imagery. International Archives of Photogrammetry and Remote Sensing, Graz, Austria, September 9-13, Vol. 34 (B3), B142- B145.

Meyer, F., and Beucher, S., 1990: Morphological Segmentation. Journal of Visual Communication and Image Representation, Vol. 11, 21-46.

Olima, W.H.A., 2001: The Dynamics and Implications of Sustaining Urban Spatial Segregation in Kenya – Experiences from Nairobi Metropolis. International Seminar on Segregation, Institute of Land Policy, Cambridge, MA, USA, July 25-28, 2001. http://www.lincolninst.edu/pubs/PubDetail.aspx?pubid=608, last visited: 17.12.2007.

Paar, G., and Pölzleitner, W., 1992: Robust disparity estimation in terrain modelling for spacecraft navigation. Proc. 11th ICPR, International Association for Pattern Recognition, 738-741.

Petrie, G., 2002: The ACSM-ASPRS Conference: A Report on the Washington Meeting, GeoInformatics, Vol. 5 No. 6, 42-43.

Raggam, H., Franke, M., Ofner, M., and Gutjahr, K., 2005: Accuracy assessment of vegetation height mapping using spaceborne IKONOS as well as aerial UltraCam stereo images. 25th EARSeL Symposium, Workshop on "3D Remote Sensing", Porto, Portugal, Jun. 10-11.

Savopol, F., Leclerc, A., Toutin Th., and Carbonneau, Y., 1994: La correction géométrique d'images satellitaires pour la Base nationale de données topographiques. Geomatica, été, Vol. 48, 193-207.

Schenk, F.A., 1996: Digital Elevation Model Extraction. In: Digital Photogrammetry: An addendum to the manual of photogrammetry. American Society for Photogrammetry and Remote Sensing. Ed. C. Greve, Bethesda, USA, ISBN: 1570830371, 247 p.

Shapiro, L.G., and Haralick, R.M., 1987: Relational matching. Applied Optics, Vol. 26, 1845-1857.

Serra, J., 1982: Image Analysis and Mathematical Morphology. Academic Press, London.

Soille, P., 2003: Morphological image analysis: principles and applications. Springer-Verlag, Berlin.

Soille, P., and Pesaresi, M., 2002: Advances in mathematical morphology applied to geosciences and remote sensing. IEEE Trans. Geosci. Remote Sensing, Vol. 40, 2042-2055.

Tao, V., and Hu, Y., 2002: 3D Reconstruction Methods Based on the Rational Function Model, Photogrammetric Engineering and Remote Sensing, Vol. 68, 705-714.

Toutin, T., 1995: Generating DEM from Stereo Images with a Photogrammetric Appraoch: Examples with VIR and SAR Data. EARSeL Journal Advnaces in Remote Sensing, Vol. 4, No. 2, 111-117.

Toutin, T. and Cheng P., 2000: Demystification of IKONOS, Earth Observation Magazine. July, Vol. 9, No. 7, 17-21.

Toutin, T., Chénier, R., and Carbonneau, Y., 2001: 3D geometric modelling of IKONOS GEO images. Proc. Joint ISPRS Workshop, “High Resolution Mapping from Space“, Hannover, Germany, Sept. 19-21, Hannover, Germany.

Toutin T., and Cheng, P., 2001: DEM with Stereo IKONOS: A reality if…. Earth Observation Magazine, Vol. 10 No. 7, 13-17.

Toutin, T., 2001: Elevation Modelling from Satellite Data. International Journal of Remote Sensing, Vol. 22, 1097-1125.

Toutin, T., Chénier, R., and Carbonneau, Y., 2002: 3D Models for High Resolution Images: Examples with QuickBird, IKONOS and EROS. Proceedings of Joint International Symposium on Geospatial Theory, Processing and Applications, Ottawa, Ontario, Canada, July 8-12.

Toutin, T., 2003a: Error tracking in IKONOS geometric processing using a 3D parametric modelling. Photogrammetric Engineering and Remote Sensing, Vol. 69, 43-51.

Toutin, T., 2003b: Block bundle adjustment of Ikonos in-track images. International Journal of Remote Sensing, Vol. 24, 851-857.

Toutin, T., 2003c: Block bundle adjustment of Landsat7 ETM^{+} images over mountainous areas. Photogrammetic Engineering and Remote Sensing, Vol. 69, 1341-1349.

Toutin, T., 2003d, Path processing and block adjustment with RADARSAT-1 SAR images. IEEE Transactions on Geoscience and Remote Sensing, Vol. 41, 2320-2328.

Toutin, T., 2004: Review Article: Geometric Processing of Remote Sensing Images: Models, Algorithms and Methods. International Journal of Remote Sensing, Vol. 25 No. 10, 1893-1924.

European Commission

EUR 23255 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen

Title: 3D-Building Height Extraction from Stereo IKONOS Data - Quantitative and Qualitative Validation of Digital Surface Models - Derivation of Building Height and Building Outlines

Author: Sandra Eckert

Luxembourg: Office for Official Publications of the European Communities

2008 – 57 pp. – 21 x29.7 cm

EUR – Scientific and Technical Research series – ISSN 1018-5593

ISBN 978-92-79-05127-2

Abstract:

This report is dealing with the digital surface model generation from VHR stereo satellite data with the focus on building height and shape extraction. The report provides a theoretical insight into orthorectification methods based on either empirical or rigorous, physical models and the theoretical aspects of digital surface model extraction. Orthorectification of stereo satellite data highly influences the accuracy of a digital surface model besides the selected matching methodology applied during the surface model extraction process. The requirement and ideal distribution of ground control points is discussed. In the final part of the report the results of four software packages, ENVI, PCI Geomatica, RSG and Leica Photogrammetric Suite, tested for urban DSM generation, are presented and described.

The orthorectification accuracy analyses were done using QuickBird and IKONOS data. For the digital surface model accuracy analyses stereo IKONOS data were mainly used. The data is commercially purchasable and is the satellite data with the highest geometric resolution that can easily be acquired as stereo datasets. Two datasets were used to perform the tests. One study area is situated in Nairobi where a variety of building types are present, from high-rise buildings to small illegal shacks. The second study area is in Graz, which was mainly chosen because a very detailed reference surface model was available.

The orthorectification accuracy results for both test areas show that the rigorous physical model performed best with an accuracy below pixel size with RMSE of 0.31m (x-direction) and 0.45m (y-direction) for the Nairobi QuickBird dataset. The empirical rational function based orthorectification achieved RMSE larger than the pixel size of 0.60m. A 1-order polynomial adjustment resulted in slightly better accuracies than a 0-order polynomial adjustment.

The orthorectification for the Graz IKONOS dataset resulted in similar accuracies: with the rigorous model RMSEs of 0.56m (x-direction) and 1.06m (y-direction) were achieved. The rational function based orthorectification with 1-order polynomial adjustment resulted in RMSEs of 0.97m and 0.63m. However, large RMSE of more than 10m for one of the used ground control points indicates that the model is not stable for the entire test area.

The physical models proved to be stable for both IKONOS and QuickBird data using few GPCs, eight for Nairobi, or a large number of GCPs, 26 for Graz. Consequently it is recommended to use the rigorous physical model for orthorectification of VHR satellite data if GCPs are available and a high geometric accuracy of the data is necessary. The RF model is still a viable alternative when high accuracy GCPs are very limited or not available. If the topography is rather flat in a dataset a 0-order polynomial adjustment with RF model orthorectification might be sufficient but a 1-order polynomial adjustment should be preferred if the terrain is rugged.

In the Nairobi test area mainly qualitative analyses and pointwise quantitative analyses were performed due to lack of reference data e.g. building heights and/or building outlines or a high-resolution digital surface model. The generated DSMs were evaluated by comparing them with reference height data taken from the internet, the GPS ground elevation data collected in the field, and data calculated by an alternative height extraction methodology. Additionally, two qualitative tests were conducted to come to a conclusion in terms of DSM quality relating to building height and shape extraction.

The five evaluation tests have shown that all tested software packages created DSMs that performed well in at least one of the tests. They all have advantages and disadvantages. Height accuracy as well as clear building shape extraction is of great importance for the use of DSMs in information extraction for settlement analysis and mapping. The highlighted tests are representing these criteria best. Judging them it can be concluded that overall the PCI and RSG software performed best. They should be favoured for DSM extraction. However, a big disadvantage of RSG is the computation time. Still, both software packages, PCI and RSG are recommended

for urban DSM extraction.

The quantitative accuracy assessment for the test area of Graz has shown that the best vertical estimation results were achieved with the software packages of LPS and PCI followed by RSG. The vertical MAE for built-up and impervious areas was 2.20m for PCI, 2.28m for LPS and 2.55m for RSG respectively. The RMSE was 3.05m, 2.96m and 3.25m respectively. Besides the vertical also the horizontal error should be considered depending on the different orthorectification methodologies applied (rigorous physical or RF based models). The shift compared with the reference DSM was between 3.06m and 3.27m. However, the qualitative, visual DSM evaluation has not confirmed the quantitative results. LPS with the best quantitative accuracy created fuzzy building outlines and contains low details in areas with smaller objects. PCI and RSG both produced DSMs with clear building outlines. They both are able to extract high details in areas with small buildings. Besides achieving the largest error in the quantitative analysis due to an erroneous mountain in the North of the test area ENVI also had problems in extracting correct multi-storey buildings outlines in the denser city area. It achieved good visual results with high details for rather small buildings.

Summarizing it can be said that the recommendations made for the Nairobi test area were confirmed by the quantitative and qualitative accuracy assessment done in the Graz test area. Both, PCI and RSG performed well or achieved at least acceptable results in both the quantitative and qualitative analysis. They both are recommended for digital surface model generation over built-up areas and settlements. Although LPS achieved the best quantitative accuracy it failed in creating DSMs with high details and clear building outlines. This is essential for building height extraction if the building outlines have to be still extracted from the data itself and are not available from cadastral offices as vector data layer. At last, ENVI showed a weak performance in creating large erroneous elevations in the test areas. It failed in achieving good quantitative results for both buildings and ground elevations. However, it should be mentioned that it successfully extracted fine structures of very small buildings in the Nairobi test area.

Two problems have to be addressed to extract building heights from stereo satellite data. First, the object height information has to be derived from the generated DSM. Two methodologies were presented to derive the object height layer: an indirect and a direct methodology. Second, the building outlines have to be delineated and extracted. A possible approach was proposed based on watershed segmentation.

The first results of the two tested methodologies are promising. A mean absolute error of 4.53m and 5.97m respectively was achieved when comparing them with reference building heights. Medium-height buildings were estimated well with an approximate error of one floor. Tall buildings are estimated with larger errors of two or more floors. These discrepancies have to be further analyzed. In case they are constant or linear the addition of an offset could be integrated into the current extraction methodologies.

Additionally, a building outline extraction approach based on watershed segmentation and preliminary results were presented. The methodology successfully detected most buildings. However, problems occur where buildings have complicated outlines. The extracted shapes of most building outlines are approximated and not representing the generally rectangular shapes of buildings. The next working steps will focus on the improvement of these approaches.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB-NA-23255-EN-C

ISBN 978-92-79-05127-2