GENERATION OF COARSE 3D MODELS OF URBAN AREAS FROM HIGH RESOLUTION STEREO SATELLITE IMAGES

Thomas Krauß, Manfred Lehner, Peter Reinartz

German Aerospace Center (DLR), Remote Sensing Technology Institute
PO Box 1116, 82230 Wessling, Germany, thomas.krauss@dlr.de

Commission I, WG I/5, ThS-3

KEY WORDS: digital terrain model, surface modelling, stereo image, high resolution, optical satellite sensors, image matching, classification, contouring, visualization, information extraction

ABSTRACT:

With the emergence of more and more satellites delivering very high resolution (VHR) imagery with ground sampling distances in the range of one meter or below the generation of three dimensional urban models directly from space may become possible. Such models are required for many applications in areas where no up-to-date detailed urban mapping exists like in developing countries. Besides the creation and updating of maps from sprawling urban settlements, such three dimensional models are also very useful for simulation and planning. For example simulations of catastrophic events like flooding, tsunamis or earth quakes rely on digital terrain models (DTM) populated with three-dimensional man made and natural objects. Using VHR satellite imagery is often the faster and cheaper alternative to acquiring aerial photos or even laser DEMs or in some times even the sole source of information for remote areas. In this paper a method for an automatic processing chain for urban modeling based on stereo images from VHR satellites is proposed. After import and preprocessing of the images a digital surface model (DSM) is derived from the stereo data. Subsequently a digital terrain model (DTM) and true ortho images are generated based on the DSM. Using a high objects mask and a vegetation mask based on the normalized difference vegetation index (NDVI) a coarse classification will be derived and suitable models for the detected objects (e.g. houses, trees, …) are selected and finally the 3D models get exported in a suitable format like VRML.

1. INTRODUCTION

The launch of WorldView-I on the 18th of September 2007 added a new member to the fleet of very high resolution (VHR) earth survey satellites. VHR satellite images with ground sampling distances (GSD) in the range of 0.5 to one meter are now available from Ikonos-2 (launch 24.09.1999, GSD pan 1 m, multispectral (MS) 4 m), QuickBird (launch 18.10.2001, GSD pan 0.6 m, MS 2.4 m), OrbView-3 (GSD pan 1 m, MS 4 m 06/2003-03/2007), and WorldView-I (GSD pan 0.5 m). Also some more systems are planned for the near future like GeoEye-1 (planned launch mid 2008, GSD pan 0.4 m, MS 1.65 m, (DigitalGlobe, 2008)), WorldView II (2009, GSD pan 0.5 m, MS 1.8 m), or the Pleiades-HR (2010, GSD 0.7 m, MS 2.8 m).

Imagery with such high resolutions allow for the first time the generation of high resolution urban models directly from space for nearly every point on earth. Throughout the following investigations Ikonos VHR stereo image pairs were used.

Already in 2001 C. S. Fraser showed that Ikonos stereo imagery has the potential for high resolution building reconstruction (Fraser et. al., 2001). A semi automatic approach can be built upon such VHR data with rather good results. However only few papers exists on fully automatic processes for extraction of urban objects from high resolution satellite data.

In contrast many approaches exist for city modelling from high resolution aerial imagery, laser scanner data, or semi automatic modelling from high resolution imagery. For a short overview or further information please refer to (Brenner, 2003) or the Ascona proceedings (referenced also in Brenner, 2003). These methods are mostly based on cadastral data, aerial images, aerial and terrestrial laser scanner data, terrestrial photographs and more additional information since the aim of these methods are often near photorealistic city models in industrial countries integrating data from several sources in often intense manual work for the urban models (CyberCity, 2008, 3D Geo, 2008). For the future Pleiades satellites (Flamanc, 2005) proposed a framework for the generation of 3D city models using a digital surface model (DSM) and a true ortho image generated from the high resolution satellite data together with 2D footprints of the buildings.

In contrast we want to propose a (mostly) automatic method for deriving 3D city models from only one single stereo scene of any area in the world. So no additional data besides the two VHR stereo images will be available since the results should also be usable for rapid mapping purposes in catastrophic events.

In this case only coarse city models with a level of detail between LoD-1 and LoD-2 will actually be needed. Following the definition in (CityGML, 2008) LoD-1 describes a block model with buildings as polygons with only a flat roof and LoD-2 describes buildings with a more detailed outside cover including more detailed roof structures and simple textures. In our case a coarse distinction between buildings with flat roofs and gabled roofs but no more roof details may be possible depending on the quality of the generated DSM.

The afterwards described processing chain is based on a VHR stereo scene. In the first step a high resolution digital surface model (DSM) has to be extracted. This step is very crucial for the quality of the result if no additional data like building footprints are available.

Most approaches for an automatic generation of 3D city models in literature are based on a high quality DSM. For example (Brenner, 2000/2003) describes fully automatic reconstruction systems based on a high quality laser DSM. Also (Gamba 2005) and (Rottensteiner, 2002) use high quality laser DSMs.

In contrast the DSMs which can be derived from optical VHR satellite imagery suffer from much errors and outliers as shown in Figure 1.

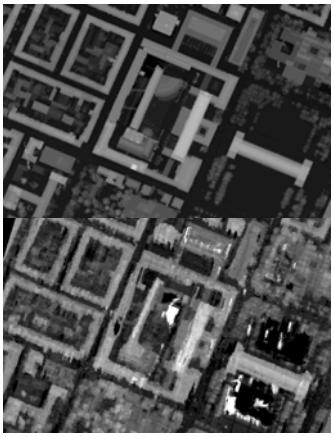

Figure 1. Left: laser DSM, right: calculated DSM from Ikonos stereo image pair (Section 600 m × 400 m from the Munich scene, area of Technical University)

Especially the estimation and extraction of building outlines will become tedious due to streaking effects, outliers, occlusions, and other artefacts in the calculated DSM. A main problem originates in the demand of a DSM with a ground resolution in the same order of magnitude as the GSD of the imagery. This requirement arose from the necessity to distinguish and characterize urban objects which requires DSMs with resolutions of about 1 m or below.

Surface models generated from aerial photographs use in contrast about 25 image pixels for one DSM pixel. Hence such DSMs are much better and can already be used for the widely proposed building extraction approaches in the literature.

2. DATA

The proposed automatic processing chain uses VHR stereo image pairs which are best acquired in the same orbit with the same illumination conditions. For the presented work two Ikonos stereo image pairs of the cities of Athens and Munich were used.

The Athens scene was acquired 2004-07-24, 9:24 GMT, with a ground resolution of 88 cm and viewing angles of –19.99° and +13.17° respectively. The scene was delivered as a level 1B image, full sensor corrected standard stereo product in epipolar geometry (Figure 2)

Figure 2. Section 600 m × 400 m from the Athens scene, left and right stereo image

Figure 3. Section 600 m × 400 m from the Munich scene (area of Technical University), left and right stereo image

The Munich scene was acquired on 2005-07-15 at 10:28 GMT with a ground resolution of 83 cm. The viewing angles of the forward and backward image were +9.25° and –4.45°. The images were available only as level 1A product, which are corrected for sensor orientation and radiometry (Figure 3) but contain no further geometric changes.

3. PROCESSING CHAIN

The proposed automatic processing chain consists of the following steps which are explained in brief in the next sections:

1 Preprocessing of the raw imagery
2 Creating the digital surface model (DSM)
3 Extracting the digital terrain model (DTM)

4 Calculating a normalized digital elevation model (nDEM)
5 Creating true orthophotos
6 Classification
7 Object extraction
8 Object modeling
9 Representing the object models through geometric primitives and exporting in a suitable 3D format

3.1 Preprocessing of the raw imagery

The Ikonos images are accompanied by rational polynomial coefficients (RPCs) describing the sensor model, orbit, and attitude data. These 80 coefficients together with 10 scale and offset parameters describe rational polynomial functions linking the geographical coordinates latitude, longitude and height above WGS84 ellipsoid with the pixel coordinates of each image (Jacobsen et al., 2005, Grodecki et al., 2004).

Unfortunately the absolute positioning of the RPCs in the case of Ikonos is only correct within a range of 10 to 50 m. Due to this in the preprocessing step a relative correlation of the two images has to be guaranteed. Therefore the two stereo images undergo an image matching process that delivers correlated points in the two images. With the knowledge of the two pixel coordinates in both images and the requirement of the same absolute height of each correlated point pair one of the RPCs can be corrected by minimizing the residuals to fit the other (Lehner et. al., 2007). In the case of Ikonos images this correction is mostly only a simple shift.

Also a pan sharpened image pair will be generated from the pan channel and the quarter resolution multispectral channels.

3.2 Creating the digital surface model (DSM)

The most crucial step in the processing chain is the generation of a rather good digital surface model from the optical VHR stereo image pair. For this task various methods where analyzed and rated for usability for such imagery. The four evaluated methods were:

- Digital line warping, "DLW" (Krauß et al., 2005)
- Semi global matching, "SGM" (Hirschmüller, 2005)
- GraphCut (Collins, 2004)
- Standard (Lehner and Gill, 1992)

In a first approach for the generation of the DSM from a stereo image pair the so called "standard" approach was analyzed. It was developed for the generation of digital surface models of images from the DLR three line scanner camera MOMS (MOMS, 1998) flown on the MIR space station. The method is based on a classical area-based matching relying on extracted interest points and an optimized region growing. In urban situations containing many steep edges and relatively large incidence angles – as used in the standard stereo products of the satellite imagery providers – only a small amount of usable 3D-points remain due to large occlusions.

So dense stereo approaches like dynamic line warping and semi-global matching were also implemented and analyzed for inclusion in the automatic processing chain. Such dense stereo methods depend however on strict epipolar geometry. A good overview of a selection of such algorithms is given on the Stereo Vision Research Page of the Middlebury College maintained by Daniel Scharstein and Richard Szeliski (Scharstein and Szeliski, 2008).

Digital line warping is based on the application of a speech recognition algorithm based on dynamic programming to coregistrated image lines in epipolar direction. Two epipolar lines of the two stereo images are correlated respectively and local distortions along the lines are calculated which lead to the local parallaxes. Due to only correlating the images line by line this method suffers from missing inter-line information which results in line streaking effects along the epipolar line.

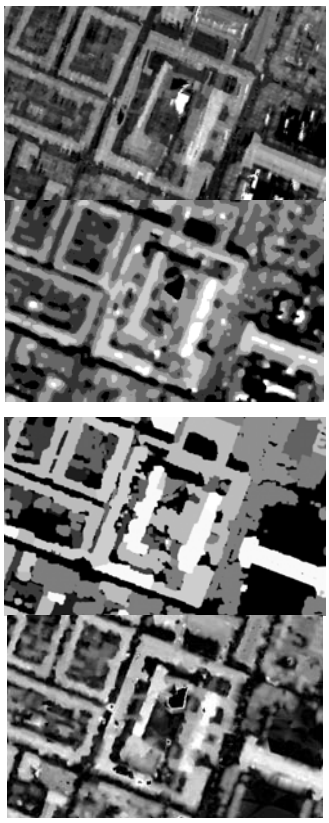

Figure 4. Top (left to right): digital surface models calculated with the methods DLW and SGM, bottom: calculated with GraphCut and Standard

The semi global matching is an extension to this approach. In this case not only two single lines of the images get compared but the energy function used by the dynamic programming integrates additional information from the whole image ("semi global"). This extension leads to much less streaking effects but also increased processing time.

The GraphCut algorithm is based on a description of the "matching space" (all 3D points of the scene) by a discrete mathematical graph with nodes at each (x,y,z) coordinate and rectangular edges connecting these nodes. The calculation of a "maximum flow" through this graph gives the correlated "minimum cut" which represents the searched surface. This method lacks in the generation of sub pixel height levels which means that a finer height resolution needs a more complex graph and hence exploding processing time.

The analysis of all algorithms with respect to a given ground truth defined by a laser DSM of the Munich scene leads to the following ranking of the methods (Pentenrieder, 2008):

Rank	Method	RMSE [m]	Visual evaluation	Time [min]
1	SGM	3,74	21	10,40
2	DLW	4,53	20	4,48
3	Standard	4,66	19	24,56
4	GraphCut	4,74	20	154,22

Table 1. Ranking of the analysed methods for generation of a digital surface model

All following investigations were carried out with the resulting digital surface model of modified versions of the two dynamic programming algorithms – digital line warping and semi global matching –, which becomes necessary due to the non-epipolar geometry of the image pairs. So the programs have to use directly the orbit and attitude information provided by the corrected RPCs to avoid an intermediate resampling step to epipolar geometry. Since the ground resolution of the satellites is in the range of one meter the generated surface model in the same resolution is rather rugged in comparison to surface models from airborne camera or lidar data.

Occluded pixels for which no height can be determined will be filled with the lowest neighbour value for visualization purposes of the DSM and marked as undefined for further processing if seen in none of the two stereo images.

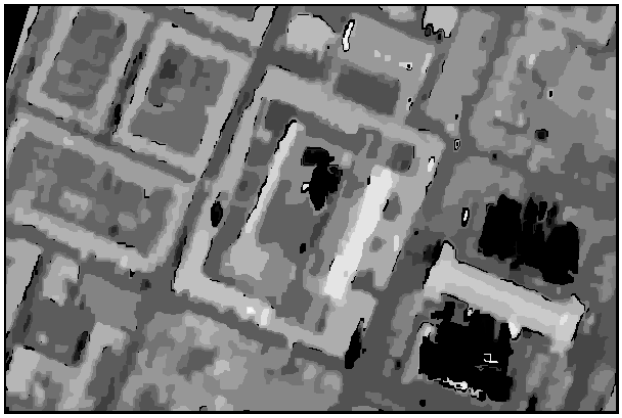

Figure 5. Digital surface model calculated for a section of 600 m × 400 m from the Munich scene using the "dynamic line warping" approach

3.3 Extracting the digital terrain model (DTM)

Using the DSM a digital terrain model describing the ground can be derived. This is accomplished by calculating a morphological erosion with a filter size of the maximum of the smallest diameter of all buildings. This results in a height image with every pixel representing the minimum height in this area around the pixel. This approach already described in (Weidner and Förstner, 1995) fails in cases of DSMs containing outliers below the real terrain. Such values will dominate the resulting DTM. So in our processing chain the morphological erosion was replaced by a median filter returning a rather low order value. After filtering an averaging using the same filter size is applied to obtain a smoother DTM. In the Munich scene shown above the DTM simply reduces to a flat plane on street level. A more sophisticated example using the Athens scene is shown in Figure 6.

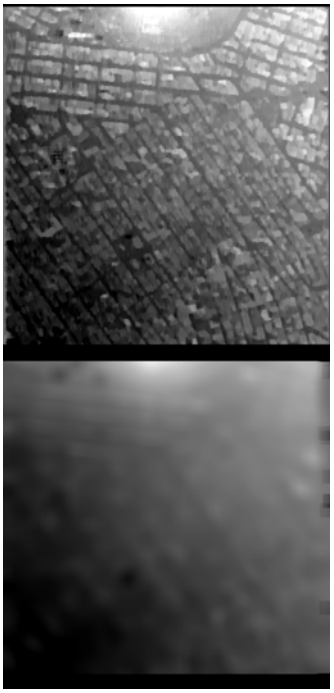

Figure 6. Sections 1000 m×1000 m from the Athens scene, left: DSM, right extracted DTM

3.4 Creating a normalized digital elevation model (nDEM)

Subtracting the DTM from the DSM gives a so called normalized digital elevation model consisting of the height of objects above the ground. In the Munich example the nDEM looks quite identical to the DSM due to the fact that the DTM is nearly flat in the shown area. In more hilly urban areas like the section of the Athens scene shown in Figure 6 the subsequent usage of an nDEM instead of the DSM becomes more important. The relation between DSM, DTM and nDEM is visualized in Figure 7.

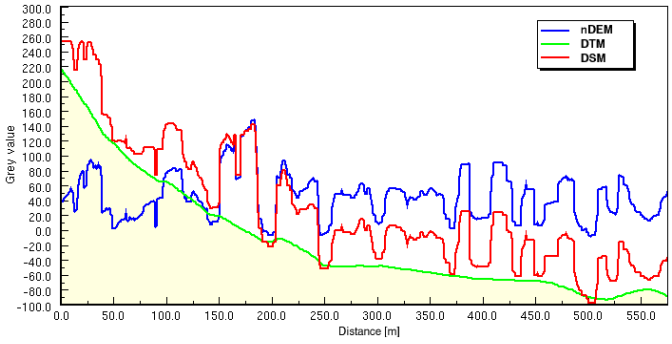

Figure 7. Profile across DSM, derived DTM and calculated nDEM for a section from the Athens scene (profile from the hill in the upper center to the center of the image; Gray-Values are arbitrary height units (parallaxes))

3.5 Creating true orthophotos

Thanks to the rather dense DSM, the RPCs from the original imagery and the pansharpened multi-spectral stereo images it is possible to derive true orthophotos. In the extracted DSM pixels occluded in both stereo images were marked as

undefined. Such positions get filled with also a value marking them as undefined in the true orthophoto. In the case an orthophoto pixel is seen from only one image of the stereo pair it is filled with this value, in the case it is visible in both stereo images it got filled with the average color of both pixels to reduce specular reflections.

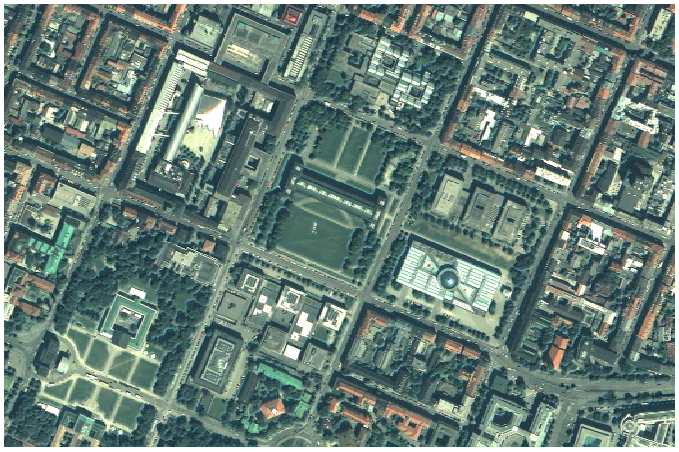

Figure 8. Pan sharpened orthophoto based on the left and right stereo image and the DSM from the Munich scene

3.6 Classification

Applying a height-threshold to the nDEM and a vegetation-threshold to the NDVI derived from the true orthophoto two binary masks are derived which allow a coarse classification into the classes: low level and no vegetation, low level with vegetation, high level and no vegetation and high level with vegetation.

The "high objects mask" is derived from the nDEM by applying a threshold of "high" (about 4 m) as shown in Figure 9.

Figure 9. High objects mask calculated from the nDEM applying a height threshold of 4 m (section 600 m × 400 m)

The "vegetation mask" is derived from the normalized difference vegetation index (NDVI) which is calculated from the red and near infrared channels of the pansharpened multispectral true orthophoto by applying a suitable vegetation-threshold (Figure 10).

$$NDVI = (NIR - Red) / (NIR + Red)$$

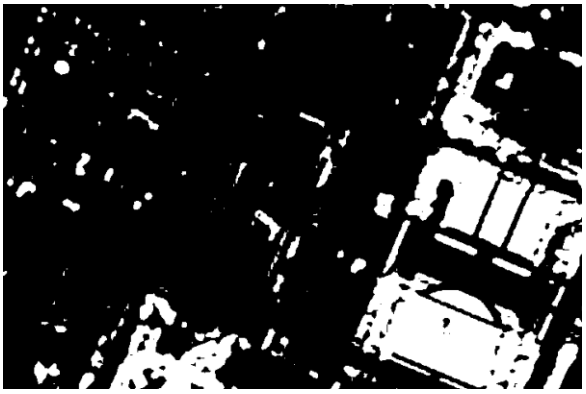

Figure 10. Vegetation mask based on the thresholded NDVI calculated from the orthophoto (600 m × 400 m)

Combining these two binary masks leads to four classes:

- low and no vegetation: streets, plain soil, . . .
- high and no vegetation: buildings, . . .
- low and vegetation: meadows, grass, . . .
- high and vegetation: trees, bushes, . . .

Figure 11 shows these classifications for the section used from the Munich scene:

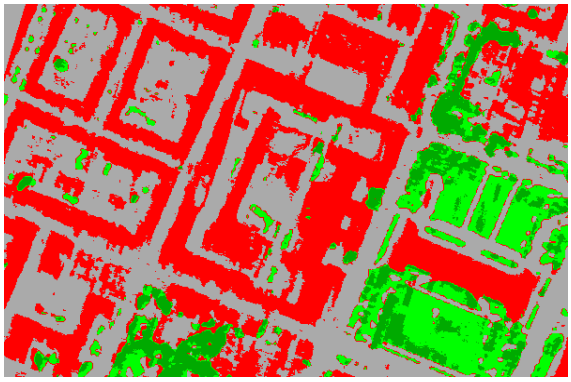

Figure 11. Classification of the Munich scene using a height mask derived from nDEM and a vegetation mask based on the NDVI from the pan-sharpened orthophotos

3.7 Object extraction

For extracting objects the nDEM and the orthophoto will be masked with one or more of the derived classes. Extracting the "high vegetation" class yields trees and bushes. The "high non vegetation" class will result mostly in man made buildings. Extracting all "low" objects will result in a ground plane. Each of these binary "one class masks" will undergo a morphological erosion to filter streaking effects and separate loosely connected parts. Afterwards the masks get separated into single connected areas – the derived objects. All of these separated objects will be fed to a process which extracts the outline and model parameters.

However this object and outline extraction step needs some more sophisticated algorithms. All approaches shown in literature for the extraction of buildings and other objects from laser DSMs do not work well for the rather rugged DSM generated by stereo matching of the VHR satellite stereo images. Some solutions presented in earlier works (Krauß et. al., 2007) will not work automatically in highly urban areas like the Athens scene and so completely different approaches have to be developed.

3.8 Object modeling

For the simple modeling of the extracted objects following base models are used:

- Model "ground" (class "low", any type of vegetation)
- Model "tree" (class "high" and "vegetation")
- Model "building" (class "high" and "no vegetation")

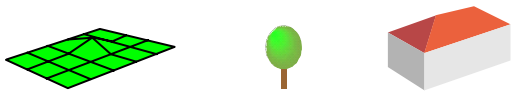

Figure 12. Simple models used

The "ground" is inserted as a height field extracted from the DTM with an optional texture directly from the true orthophoto.

"Trees" are described by a crown diameter and a treetop height extracted from the classification and the DSM respectively. "Buildings" are represented as prismatic models. In the future the prismatic models will be split to cuboids with optionally parametric gabled roofs.

3.9 Representing the object models through geometric primitives and exporting in a suitable 3D format

The coarse models will be represented by geometric primitives. A height field derived from the DTM for "ground" (one for the full scene, textured from the true orthophoto), an ellipsoid supported by a cylinder for trees and rectangular vertical walls following the extracted circumference and a horizontal polygonal roof in the first version. A texture may be extracted from the original images by projecting the resulting polygons backward using the RPCs. The optionally textured geometric primitives have to be exported into a suitable 3D vector format.

The automatic export step of the chain already works well. Required inputs are simply the DSM, the true ortho photo matching the DSM, and the objects in form of 2D vector outlines. These outlines represent the two elevated classes: trees will be marked by circles, polygons represent buildings or parts of buildings with the same roof slope. In the Athens example these outlines were generated manually and only flat roofed buildings occurred in the scene (see Figure 13).

In the export step automatically the height of the objects is extracted from the outlines and the DSM. A totally new digital terrain model will be created from the DSM by cutting out all elevated objects marked by the 2D outlines and consecutively interpolation and smoothing. The ground object will also be textured by the true ortho photo.

Figure 13. Orthorectified section from the Athens scene (UTM projection) with manually marked trees (green circles) and building outlines (orange polygons)

Figure 14 shows the automatically generated 3D model with the textured ground and elevated tree- and building-objects from the lower left quarter of the Athens scene in Figure 13 using a VRML viewer.

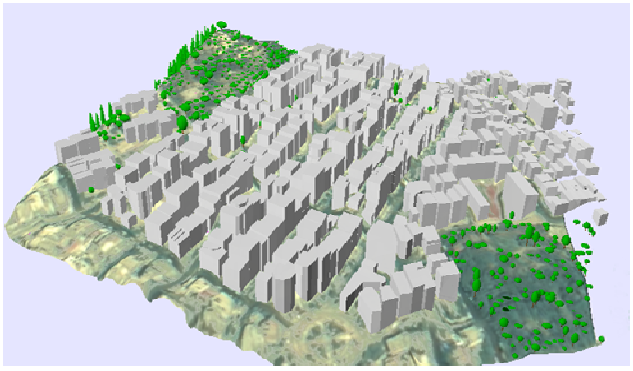

Figure 14. 3D view automatically generated from the DSM and the manually marked trees and building footprints from the Athens scene, size 500 m × 500 m

4. SUMMARY AND OUTLOOK

In this paper a first version of a processing chain for the automatic extraction of three-dimensional city models directly from high-resolution stereo satellite images is described. The chain elements are already implemented but some steps require further optimization. For example the DSM generation needs some major improvement. Also the methods developed in previous works for building extraction fail to work in some complex urban areas. So the main future work will focus on

the extraction of trees and man made objects from the best possible DSM and VHR satellite imagery. But the results gained up to now from each step are encouraging enough to follow the path and refine every step of the chain to receive a new fully automatic system for generating coarse three-dimensional urban models from stereo satellite imagery in a short time.

REFERENCES

3D Geo: http://www.landxplorer.net/ (accessed 04/2008).

Birchfield, S. and Tomasi, C., 1998: Depth discontinuities by pixel-to-pixel stereo. Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India, pp. 1073–1080.

Brenner, C., 2000: Towards fully automatic generation of city models, ISPRS, Vol. 33, Amsterdam, 2000

Brenner, C., 2003: Building reconstruction from laser scanning and images. Proc. of the ITC Earth Observation Science Department Workshop on Data Quality in Earth Observation Techniques, Enschede, The Netherlands, November 2003

CityGML: https://portal.opengeospatial.org/files/index.php?artifact_id=26639 (accessed 04/2008)

Collins T., 2004: Graph Cut Matching In Computer Vision, University of Edinburgh, February 2004.

CyberCity, 2008: Zürich, vormals http://www.cybercity.tv/. (accessed 07/2007), inzwischen in Liquidation (http://www.moneyhouse.ch/u/cybercity_ag_in_liquidation_CH-020.3.023.691-6.htm, acc. 04/2008)

DigitalGlobe: http://www.digitalglobe.com/about/imaging.shtml. (accessed 04/2008).

Flamanc, D., Maillet, G., 2005: Evaluation of 3D city model production from Pleiades-HR satellite images and 2D ground maps. ISPRS Vol. 36 (8/W27), Tempe, AZ, USA, 2005

Förstner, W. and Gülch, E., 1987: A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: ISPRS Intercommission Workshop, Interlaken.

Fraser, C.S., Baltsavias, E., Gruen, A., 2001: 3D building reconstruction from high-resolution Ikonos stereo imagery, in Automatic Extraction of Man-Made Objects from Aerial and Space Images, publisher A A Balkema (Lisse)

Gamba, P. Dell'Acqua, F., Cesari, M., 2005: Three-dimensional object recognition in LIDAR data using a planar patch approach. ISPRS Vol. 36 (8/W27), Tempe, AZ, USA, 2005

Grodecki, J., Dial, G. and Lutes, J., 2004: Mathematical model for 3D feature extraction from multiple satellite images described by RPCs. In: ASPRS Annual Conference Proceedings, Denver, Colorado.

Hirschmüller, H., 2005: Accurate and efficient stereo processing by semi-global matching and mutual information. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Jacobsen, K., Büyüksalih, G. and Topan, H., 2005: Geometric models for the orientation of high resolution optical satellite sensors. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36 (1/W3). ISPRS Workshop, Hannover.

Krauß, T., Reinartz, P., Lehner, M., Schroeder, M. and Stilla, U., 2005: DEM generation from very high resolution stereo satellite data in urban areas using dynamic programming. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36 (1/W3). ISPRS Workshop, Hannover.

Krauß, T., Reinartz, P., Stilla, U., 2007: Extracting Orthogonal Building Objects in Urban Areas from High Resolution Stereo Satellite Image Pairs. ISPRS, 36 (3/W49A), PIA 2007, Munich

Lehner, M. and Gill, R., 1992: Semi-automatic derivation of digital elevation models from stereoscopic 3-line scanner data. ISPRS, 29 (B4), pp. 68–75.

Lehner, M., Müller, Rupert, Reinartz, P., Schroeder, M., 2007: Stereo evaluation of Cartosat-1 data for French and Catalonian test sites, Proc. of the ISPRS Workshop 2007 High Resolution Earth Imaging for Geospatial Information, Hanover, Germany, May 29 – June 1

MOMS, 1998: http://www.nz.dlr.de/moms2p/ (accessed 04/2008)

Pentenrieder, C., 2008: Analyse und Vergleich von 3D-Stereo-Verfahren für hochauflösende Satellitenbilder. Diploma thesis, Hochschule für Angewandte Wissenschaften – FH München, Fakultät Geoinformationswesen, April 2008

Otto, G. and Chau, T., 1989: Region growing algorithm for matching of terrain images. Image and vision computing (7) 2, pp. 83–94.

Rottensteiner, F., Briese, Ch., 2002: A new method for building extraction in urban areas from high-resolution LIDAR data. ISPRS Vol. 36 (8/W27), Tempe, AZ, USA, 2005

Scharstein, D. and Szeliski, R. Middlebury stereo vision page: http://vision.middlebury.edu/stereo/. (accessed 04/2008).

Weidner, U., Förstner, W., 1995: Towards automatic building extraction from high resolution digital elevation medels. ISPRS J. 50 (4), 38–49.

SpaceImaging/GeoEye: http://www.geoeye.com/ (accessed 04/2008).

ACKNOWLEDGEMENTS

The authors thank cordially Christian Pentenrieder for the comparison, analysis, and evaluation of the different DSM generation methods and Martin Brandt for the tedious job of

manual outlining the buildings and trees in the above shown part of the Athens scene.