MODELING 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

IPIworkshop, Hannover 29.05.–01.06.2007
Land Cover and Change Detection II
Commission III, WG III/4

KEY WORDS: VHR data, stereo images, digital surface models, classification, urban modelling

ABSTRACT:

Models of urban areas are an important input for many applications in the field of urban monitoring. Besides the creation and updating of maps from sprawling urban settlements the models are also used for simulation and planning in case of catastrophic events like flooding, tsunamis or earth quakes. With the availability of very high resolution (VHR) satellite data investigations of large urban areas regarding their three dimensional shape can be performed fast and relatively cheap in comparison to aerial photography especially for cities in developing countries. Most of the actual methods used for the generation of city models depend on a large amount of interactive work. A method for automatic derivation of – in a first step – coarse models of urban structure is therefore of great use. In this paper the methods used for such an automatic modeling is presented. The methods are based on stereo images from VHR satellite stereo imagery provided by satellite sensors like IKONOS or QuickBird. In a first step a digital surface model (DSM) is derived from the stereo data. Subsequently a digital terrain model (DTM) is generated from the DSM. Orthoimages are created in parallel using the derived DSM. Based on the local height differences between DSM and DTM and the normalized difference vegetation index (NDVI) a coarse classification of the orthoimages is performed. The classification divides the images in four classes: low level and no vegetation, low level with vegetation, higher level and no vegetation and higher level with vegetation. Based upon this classification object models are selected and object parameters are adapted to create an object-based representation of the satellite image scene. The used processing steps are evaluated and the results for IKONOS stereo pairs of the cities of Athens and Munich are presented and discussed.

1 INTRODUCTION

All around the world – especially in developing countries – urban areas grow in rapid speed. E.g. Mexico City grows every day by about 800 people due to migration or birth. This is an additional 300 000 per year on top of already 21.5 million inhabitants. Keeping pace with such a rapid growth of urban areas in gaining data for planning or in case of catastrophes is extremely costly with classical methods. For this the modeling of cities fully automatically directly from relative cheap satellite images will be a great help.

Beside planning and aid in catastrophic events many companies, e.g. from the telecommunication sector, rely on high resolution three-dimensional city models for determining antenna positions and wave propagation. Normally these models are created by using high resolution aerial image data or data from laser scanning and further information like cadastral data. Now from stereo satellite data of the highest resolution class the generation of coarse city models becomes possible.

In this work a first approach for the implementation of an automatic processing chain for the modeling of urban areas from very high resolution satellite image stereo pairs is suggested. In a first step a digital surface model (DSM) has to be extracted from the stereo pair. To achieve this, several different methods are tested: intensity based matching, dynamic line warping and the so called semi-global matching technique. From the DSM a ground model, a digital terrain model (DTM) will be extracted. Using the derived ortho images for a vegetation mask and the DSM and DTM for a mask of high objects the imaged area can be classified in a simple manner. Based on this coarse classification models for buildings and high vegetation can be derived, textured and exported to simple 3D-models.

2 STATE OF THE ART

Currently exist many methods for city modeling. These methods are mostly based on cadastral data, aerial images, aerial and terrestrial laser scanner data, terrestrial photographs and much more. These models integrate data from several of these sources in mostly intense manual work for the urban models [CyberCity, 2007, 3D Geo, 2007].

For example CyberCity uses a semi-automatic extraction of 3D point clouds with a photogrammetric workstation. From these points the roofs are generated and walls are built by projection of the roofs to the digital terrain model or directly from cadastral data. The textures of roofs and walls are finally extracted by projection of the models into the aerial photos. As stated by CyberCity a good operator is able to extract up to 500 roof parts per day.

In our approach using only stereo satellite images a big limitation is the missing availability of several data sources. Subject of this work is to develop a system, which creates urban models only from a minimum of two high-resolution stereo satellite images. Such images are provided at the moment, e.g., by Space Imaging (Ikonos, [SpaceImaging/GeoEye, 2007]) with a ground resolution of about 1 meter panchromatic and 4 m multispectral or Digital Globe (QuickBird, 60 cm panchromatic and 2.4 m multispectral).

In the near future many more high-resolution optical satellites are due to go into operation like WorldView I (2007) and II (2008) offering half-meter panchromatic and 1.4 to 1.8 m multispectral resolution ([DigitalGlobe, 2007]).

3 DATA

The method relies on stereo image pairs which are best acquired in the same orbit with same illumination conditions, two distinct viewing angles and known internal and external orientation of the satellite (e.g. orbital positions and look angles or RPCs).

The following investigations were done mainly with two distinct Ikonos stereo image pairs:

- Athens scene: acquired 2004-07-24, 9:24 GMT, ground resolution 88 cm, viewing angles –19.99° and +13.17°, level 1B image: full sensor corrected standard stereo product in epipolar geometry (Figure 1)
- Munich scene: acquired 2005-07-15, 10:28 GMT, ground resolution 83 cm, viewing angles +9.25° and –4.45°, level 1A image: only corrected for sensor orientation and radiometry (Figure 2)

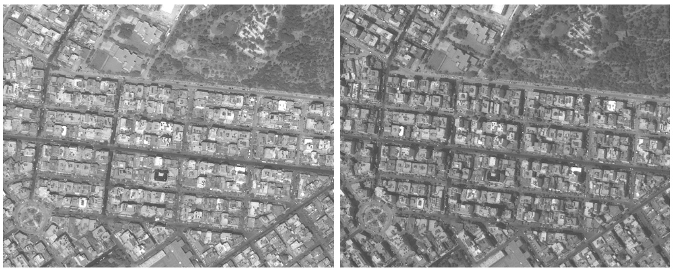

Figure 1. Section 500 m × 400 m from the Athens scene, top: left stereo image, bottom: right stereo image

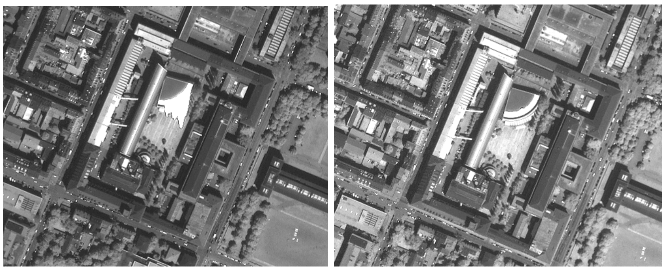

Figure 2. Section 500 m × 400 m from the Munich scene (area of technical university), top: left stereo image, bottom: right stereo image

4 PROCESSING CHAIN

The suggested processing chain consists of the following steps which are explained in detail in the following sections:

- Preprocessing of the raw imagery
- Creating a high-resolution digital surface model (DSM)
- Deriving of a digital terrain model (DTM) from the surface model (DSM)
- Calculating pansharpened images
- Projecting these pansharpened images to the DSM and creating high-resolution true ortho photos
- Classifying these ortho images to simple base classes
- Extracting distinct objects
- Selecting applicable models for all extracted objects
- Fitting model parameters for these objects
- Representing the object models through geometric primitives
- Texturing the geometric primitives
- Exporting geometric primitives into a suitable 3D-file-format

4.1 Preprocessing of the raw imagery

The preprocessing step includes the import of the images and the injection of the meta data and the rational polynomial coefficients (RPCs) which accompany high-resolution Ikonos satellite imagery for every image. These coefficients are used to transform the geographical coordinates longitude *X*, latitude *Y* and ellipsoid height *Z* to image coordinates (*x*,*y*) by division of two polynoms with 20 coefficients each [Jacobsen et al., 2005, Grodecki et al., 2004]. These geometric calculations are used throughout the generation of the DSM and further for absolute georeferencing and ortho image generation. In a further preprocessing step the multispectral channels with much coarser resolution (one pixel is about four meter) than the panchromatic channel are pansharpened.

4.2 Creating the digital surface model (DSM)

The first complex processing step is the generation of a digital surface model from the image stereo pair. For creating a “depth map” from two stereo images a broad variety of methods exists. All methods use a kind of matching of features found in the first image to corresponding features in the second image. The relative displacement of these features give the disparity vectors and so a global depth map of the scene. Two different approaches can be distinguished: Area based matching compares small windows between the images. Feature based matching compares extracted objects like lines or areas between the images.

Analyzing the different approaches in respect to the given imagery (“high-resolution” of about 1 m, steep walls, many occlusions, most features in size of resolution) leads to the following results:

- Pixel-, area-, intensity-based methods:
  - need a small correlation window to find correlating areas in the stereo image pair
  - steep features as common in urban areas smooth out on the size of this window
  - problems occur with occlusions that are common in urban scenes
- Object-, line-based methods:
  - “high-resolution” satellite imagery has a too coarse resolution for line extraction

- most of the features, which can be found in urban images, are of the size of the resolution
- so lines and features found in one image mostly have no corresponding features in the other image

A third approach besides pure area based and pure object based matching has to be used which depends not on large features and needs only small windows of about one pixel. If the geometry of the images is fairly good, known pixels in one of the images can be projected to absolute object coordinates by using the RPCs and a given height. So projecting corresponding pixels from both images with the same given height fit to the same absolute object coordinates.

This approach comes to a simple solution if epipolar images can be used. A good overview of a selection of methods applicable for such simple epipolar geometry can be found on the Stereo Vision Research Page of the Middlebury College maintained from Daniel Scharstein, Middlebury College and Richard Szeliski, Microsoft Research [Scharstein and Szeliski, 2007].

Besides these an existing intensity based method and another epipolar based method developed at the German Aerospace Center (DLR) by [Lehner and Gill, 1992] and [Krauß et al., 2005] where investigated.

In the first (classical) method interest points are searched via the usage of the Förstner interest operator [Förstner and Gülch, 1987] in one image of the stereo pair. Based on a set of six conjugate points searched manually in images scaled down in an image pyramid by a factor of 32 an affine transformation is estimated. The interest points found in the first image are transformed via this transformation to the second image of the stereo pair and an area based matching of image windows around these points is performed. Subpixel refinement via local last square fit, propagating up the image pyramid and performing forward-backward-matching delivers a good and stable set of conjugate points in the full scale stereo pair.

Applying a region growing [Otto and Chau, 1989] gives a dense quantity of points and correlated disparities. Transforming this information to object space using ephemeris and pointing data and applying a triangulation to fill a grid delivers the digital surface model.

The second approach named “dynamic line warping” was developed applying a method known in speech recognition as “dynamic time warping” line by line for stereo images. This algorithm warps one linear array of values onto another. In speech recognition these arrays are voice input and recorded speech that are compared. In stereo image matching the two arrays are the corresponding lines of the two images. This approach requires in contrast to the first approach epipolar images.

The two input arrays (image scan lines) are correlated using a dynamic programming approach. The required cost function for image matching may be a pixel-by-pixel gray value distance or the cost function used by Birchfield and Tomasi [Birchfield and Tomasi, 1998]. Based on this distance a cost matrix can be calculated. Using dynamic programming a “distance” between the two linear input arrays can be calculated. Beneath this distance (the bottom rightmost value of the resulting matrix; a measure for the needed warping operations) also the minimal path – the local distortions between the two scan lines – can be retrieved. This path yields the disparity and the parallaxes at each point of the scan line. Using the RPCs of the satellite images object coordinates can be retrieved from these parallaxes.

The advantage of dynamic line warping in respect to an area based matching approach is that the used “window” can be as small as only one pixel and so no smearing at steep edges should occur. But disadvantages of this method are the requirement of epipolar images, streaking effects occur since there is no interconnection of the epipolar lines and also errors occur in water and grass areas. One approach for avoiding the streaking between neighboring epipolar lines is the multidirectional approach „semi-global matching“ described by [Hirschmüller, 2005].

In this similar dynamic programming approach the summation occurs not only epipolar line for epipolar line but parallel for all epipolar lines up to a given maximum disparity. The disparity matrix is in this case a three-dimensional matrix of size image width × image height × ±maximum disparity. Beside this the used cost function can be calculated after Birchfield/Tomasi or using mutual information.

The three investigated methods for the calculation of the DSM are in summary:

- Classical, area based hierarchical matching (Figure 3)
- Matching epipolar lines using dynamic line warping (Figure 4)
- Matching epipolar lines using semi-global matching (Figure 5)

The results can be enhanced by applying a small median filter. So sharp edges are retained but outliers and parts of extreme streaking effects are filtered out.

Also the cost function can be expanded from only one pixel to a small window of size *–w/2* … *w/2* which creates better results but introduces on the other hand a smoothing effect on steep ridges. In all following calculations a window and median size of 3 pixels were used.

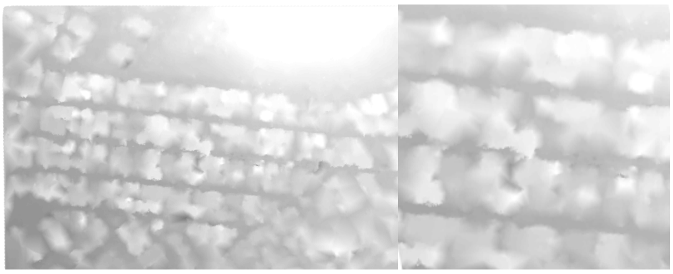

Figure 3. DSM generated by classical area based matching (Athens, all sections 640 m × 400 m and all subsections 250 m × 250 m)

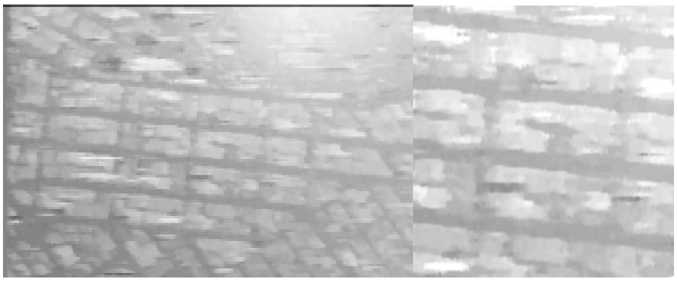

Figure 4. DSM generated by dynamical line warping

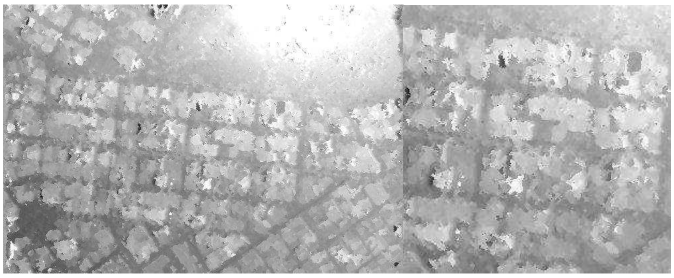

Figure 5. DSM generated by semi-global matching

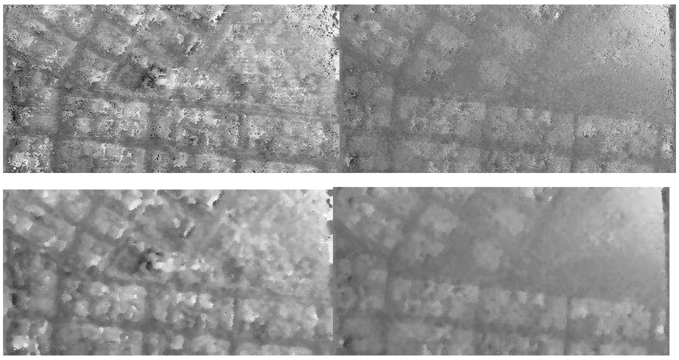

Figure 6. DSMs from Athens scene generated by semi global matching, left with intensity-based cost function (Birchfield-Tomasi), right with mutual information, upper with median of size 1, lower with median size of 5

Using mutual information in the calculation of costs (semi-global matching) instead of simple gray value differences or Birchfield/Tomasi also gives some remarkable improvements as shown in Figure 6 but increases the computing time significant due to needed iteration steps.

4.3 Extracting the digital terrain model

Based on the calculated DSM a digital terrain model (DTM) can be derived. For this first a maximum of the smallest diameter of all buildings has to be estimated. For a normal urban area this value is in general about 20 m wide since every room should have at least one window. In industrial areas the minimum diameters of buildings can be much larger so lengths up to 100 m may be necessary.

Applying a morphological erosion with a filter size of this diameter to the DSM results in a height image with every pixel representing the minimum height in this area around the pixel. For calculating the DTM in contrast to [Weidner and Förster, 1995] in reality a median filter will be applied instead of the morphological erosion, which does not return the middle height value but rather a low value. If the lowest value will be returned it will be the same as the morphological erosion but using a value of about 1 to 5 % avoid the domination of the generated DTM by single outliers from the calculated DSM. After filtering an averaging using the same filter size is applied to obtain a smoother DTM.

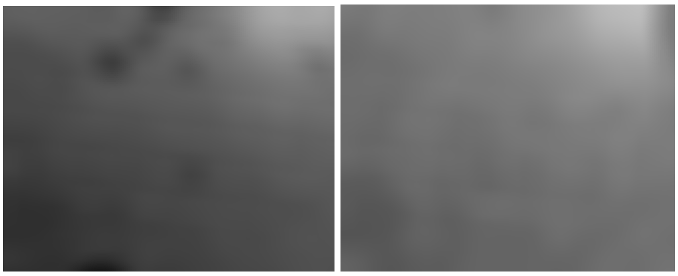

Figure 7. DTMs calculated from the DSMs resulting from the digital line warping (left, some DSM outliers propagate to the DTM) and semi global matching (right)

4.4 Creating true ortho images

Using the dense DSM, the RPCs and the pansharpened multi-spectral stereo images true ortho images can be derived. In contrast to simple ortho images – which use generally coarser DEMs – in true ortho images also buildings will be orthocorrected and not only squeezed onto ground.

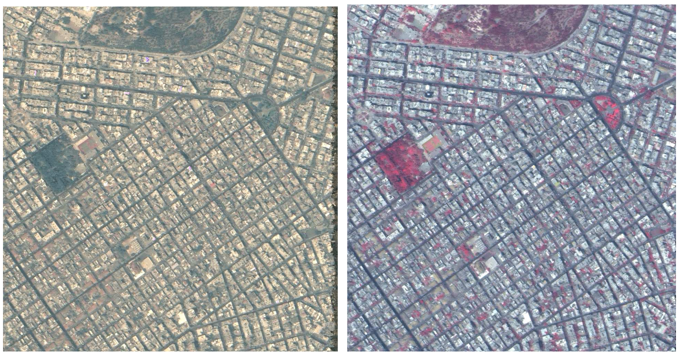

Figure 8. Orthoimage Athens scene, left: channels red, green, blue; right: channels near infrared, red, green

Figure 8 show the resulting ortho images from the Athens scene as RGB- and IRG-image respectively.

4.5 Classification

Based on the pansharpened true ortho images and the calculated DSM and DTM a coarse four class classification can be calculated from two binary masks:

- The “high objects mask” calculated as difference of DTM and DSM and applying a height-threshold as shown in Figure 9.
- The “vegetation mask” from the NDVI (normalized difference vegetation index, NDVI = (NIR – Red) / (NIR + Red)) of the pansharpened multispectral true ortho image with a vegetation-threshold (Figure 10).

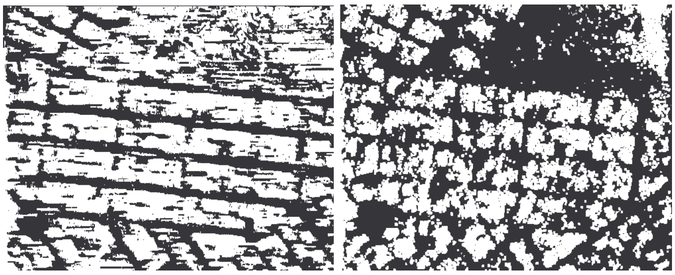

Figure 9. Height masks calculated from the line-warping-DSM (left) and the semi-global-DSM (right) (all images 500 m × 400 m)

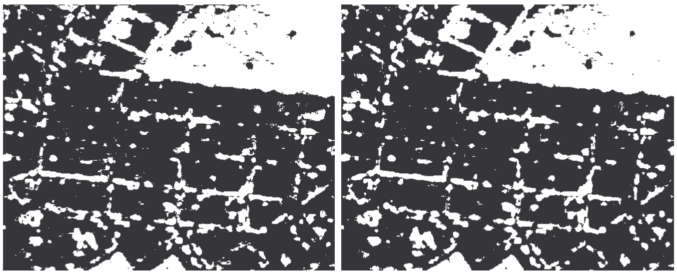

Figure 10. Vegetation masks calculated from the ortho image based on the line-warping DSM (left) and the semi-global DSM (right) (all 500 m × 400 m)

Combining the 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 two DSM generation methods respectively.

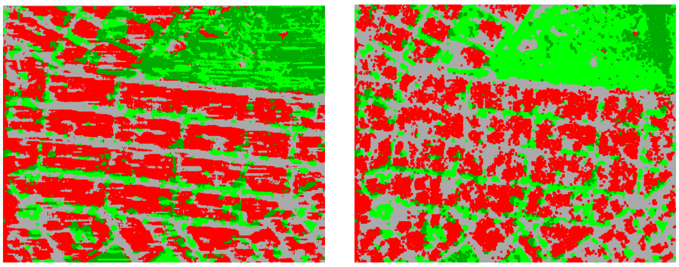

Figure 11. Classification of the Athens scene using the DSM and DTM generated by dynamic line warping (left) and semi-global matching (right)

4.6 Object extraction

Based on the classification objects can be extracted. Extracting the "high vegetation" class yields trees and bushes. The "high non vegetation" class will result mostly in man made buildings. This approach doesn't work in case of steep rocks not covered by vegetation or buildings with roof top gardens. The method will be shown below at a section from the Munich test scene (Figure 12, Technical University of Munich with four identified objects).

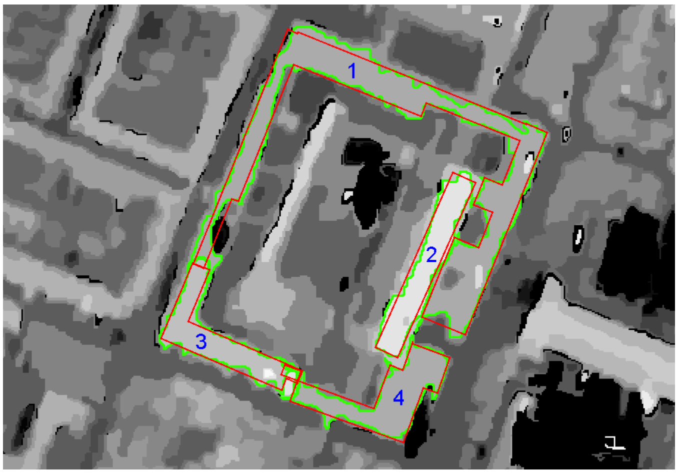

Figure 12. Four selected outlines for following investigations in Munich scene (image size: 500 m × 400 m)

After masking the DSM with the class "high non vegetation" neighboring heights are combined to height classes. Extracting the border of these height classes results in the outlines of the desired object.

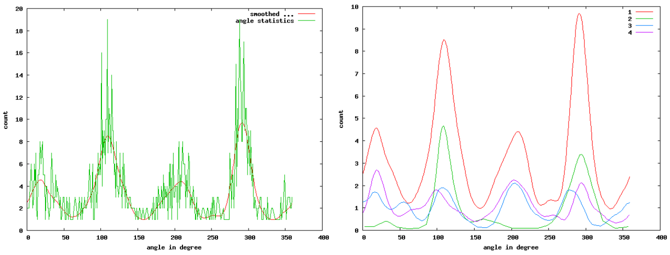

Figure 13. Statistics of measured angles for first polygon (up: count, right: angle in degree), right: smoothed distribution for all 4 borders

These borders are analyzed by creating a statistics of occurrence for the direction of each border element. Figure 13 shows this statistics for border "1" (green) from Figure 12 and the smoothed distribution for all four borders. The found maxima in this distribution will be used as main directions. For polygon 1 maxima at 22°, 112°, 202° and 292° can be located. After filtering and grouping angles to these found main directions (Figure 14) in a last step consecutive line segments (Figure 15) are intersected and the closed red polygons shown in Figure 12 above are generated.

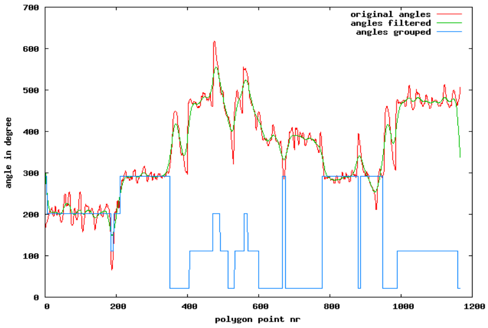

Figure 14. Grouping angles of border elements together to found main directions

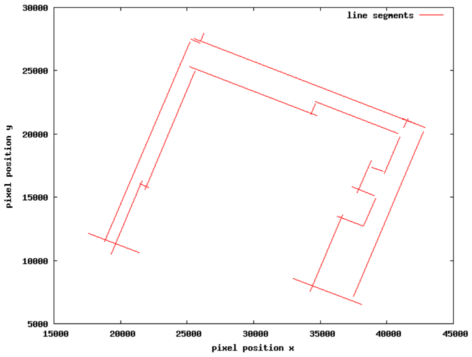

Figure 15. Derived line segments of polygon 1 after grouping of angles

4.7 Modeling

For the subsequent 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”)

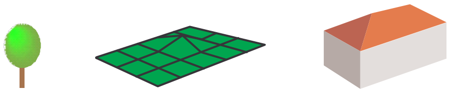

Figure 16. Simple models used

The “ground” is inserted as a height field extracted from the DTM with an optional texture directly from the true ortho image. Trees are described by a crown diameter and a treetop height extracted from classification and the DSM respectively. A surrounding rectangular polygon describing a building is extracted as described above. In a first version a coarse modeling can be done by simply generating a polygonal model with the polygon as circumference and a flat top with an average height extracted from the DSM. The accuracy and resolution of the derived DSM is in most cases not sufficient for an detailed modeling of the roof shape. In a more detailed version the polygonal circumference has to be divided into rectangles. An averaged section across these rectangles may give a hint if it's a gabled or a flat roof and lead to a more detailed model.

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

The coarse models will be represented through geometric primitives. A height field derived from the DTM for ground (one for full scene if textured from the true ortho image), a 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 can 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. Figure 17 shows the area around the technical university from the Munich scene as shown by a VRML viewer.

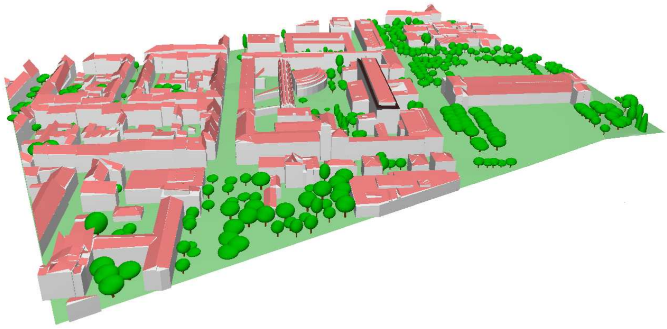

Figure 17. Simple 3D view generated from the Munich scene, size 640 m × 400 m, center: Technical University of Munich, right Old Pinacotheca

5 SUMMARY AND OUTLOOK

In this paper a processing chain for the automatic extraction of three-dimensional city models directly from high-resolution stereo satellite images is suggested. The processing chain is still in development. So the DSMs generated are not satisfying enough up to now. Also the automatic extraction of objects from the classified images exists only in a first evaluation version. The texturing is still missing up to now. But the results gained from each step are encouraging enough to follow the path and refine every step of the chain to receive a newly 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 03/2007).

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.

CyberCity. http://www.cybercity.tv/. (accessed 03/2007).

DigitalGlobe. http://www.digitalglobe.com/about/imaging.shtml. (accessed 03/2007).

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.

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.

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.

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

Scharstein, D. and Szeliski, R. Middlebury stereo vision page. http://cat.middlebury.edu/stereo. (accessed 03/2007).

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 03/2007).