SEMI-AUTOMATIC CITY MODEL EXTRACTION FROM TRI-STEREOSCOPIC VHR SATELLITE IMAGERY

F. Tack [a,*], R. Goossens [a], G. Buyuksalih [b]

[a] Dept. of Geography, Ghent University, Krijgslaan 281, 9000 Ghent, Belgium – (f.tack, rudi.goossens)@ugent.be
[b] IMP-Bimtas, 34430 Beyoglu, Istanbul, Turkey – gbuyuksalih@yahoo.com

KEY WORDS: Photogrammetry, City DSM generation, Tri-stereoscopy, Ikonos, Urban

ABSTRACT:

In this paper a methodology and results of semi-automatic city DSM extraction from an Ikonos triplet, is introduced. Built-up areas are known as being complex for photogrammetric purposes, mainly because of the steep changes in elevation caused by buildings and urban features. To make surface model extraction more robust and to cope with the specific problems of height displacement, concealed areas and shadow, a multi-image based approach is followed. For the VHR tri-stereoscopic study an area extending from the centre of Istanbul to the urban fringe is chosen. Research concentrates on the development of methods to optimize the extraction of a surface model from the bundled Ikonos triplet over an urban area, without manual plotting of buildings. Optimal methods need to be found to improve the radiometry and geometric alignment of the multi-temporal imagery, to optimize the semi-automatical derivation of DSMs from an urban environment and to enhance the quality of the resulting surface model and especially to reduce smoothing effects by applying spatial filters.

1. INTRODUCTION

The high level of detail and geometric accuracy of very high resolution satellite data such as Ikonos imagery, has made this kind of imagery suitable for DSM generation at feature level of urban environments. Due to the photogrammetric complexity of urban areas, quite some research is done to cope with the specific problems of urban surface model generation from standard stereopairs. As a multi-image based approach can make the 3D modelling more robust, a methodology and results of semi-automatic DSM production from an Ikonos triplet over an urban area, is highlighted in this paper. From a theoretical point of view the redundancy of a third image should lead to a more reliable photogrammetric processing. Only a few investigations have been published dealing with the concerning subject. Research published in (Baltsavias et al., 2006) and (Raggam, 2006) can be referred to.

2. COMPLEXITY OF URBAN AREAS

A Digital Surface Model is a digital representation of the terrain and topographic object height in a grid structure. Interpolation of the discrete height values is needed to approximate the continuity of the ground surface. Urban environments are experienced as complex for 3D modelling purposes because of the steep changes in elevation and the discrepancy between the smoothness of the ground surface and abrupt discontinuities caused by buildings and other urban features. Without manual plotting or spatial filter techniques it is difficult to reconstruct vertical walls out of VHR satellite imagery. An interpolation technique creates a smoothed surface and causes individual buildings will have a shape of a bell instead of the rectangular geometry (Jacobsen, 2006). A second consequence of steep changes in elevation is the occurrence of shadow and concealed areas. Due to the convergent viewing angle of VHR sensors like Ikonos, terrain features with certain height above the surface are geometrically displaced in the imagery, leading to dissimilarities between the stereo images. By this distortion of its true position, parts of the ground surface can be hidden in the satellite image. These are so-called occluded areas. Shadow areas, which have poor contrast, and occlusion areas lead to mismatches during the image matching algorithm and subsequently to errors in the resulting surface model.

3. DATA SET AND STUDY AREA

Next to along track stereo pairs, it is possible to create stereo couples out of images from the same area but taken from a different orbit at a different date. These are so-called across track stereo pairs. This approach to form couples has some disadvantages. The most important ones are radiometric differences and changes of the ground surface due to the time gap between acquisition of the imagery. A triplet is constructed out of an along track Ikonos stereo pair taken in March 2002 and a third image taken in May 2005. The third image can be considered as a nadir image. Selection criteria for the near vertical image were multiple: overlap with stereo couple, cloud-free acquisition, minimal time interval and optimal stereo constellation. Despite the big time interval, the 2005 Ikonos image was chosen to be the most optimal candidate.

The Ikonos STEREO product imagery, which comprises of a forward and backward image acquisition and the GEO Ortho Kit 2005 image are panchromatic, resampled to a spatial resolution of 1 m by the image provider and provided with the Rational Polynomial Coefficients (RPC) camera model file. Further characteristics of each image of the triplet can be found in table 1.

---

[*] Corresponding author

Image ID	Acquisition date	Elevation angle	Collection azimuth	Sun elevation angle
A (Forward)	1/03/2002	67.59°	1.6°	39.1°
B (Backward)	1/03/2002	75.59°	214.1°	39.1°
C (Nadir)	16/05/2005	80.93°	23.5°	65.5°

Table 1. Characteristics of the three VHR satellite images acquired over the study field.

Part of the mega city Istanbul, Turkey is chosen as test field for the project, mainly because it is a city characterized by an intense urban growth. The city is very compact and concentrated along the Bosphorus strait. The high resolution test area covers the overlapping area between the Ikonos 2002 stereo pair and the 2005 image and covers an area of approximately 60 km², containing Istanbul's historic peninsula and going up to the north to the urban fringe. It concerns a densely built-up area with a height range of 220 m with the lowest point at sea level and geo-morphologically characterized by a hilly landscape.

4. SURFACE MODEL GENERATION

In following subsections, the successive steps of the applied methodology for city surface model generation, based on (tri)stereoscopic VHR satellite imagery, are elucidated. The emphasis is especially laid on those phases were research is done to cope with the complexity of an urban environment.

4.1 Tri-stereoscopic approach

Instead of the standard stereo mapping with two images a tri-stereoscopic approach is followed. Generation of a DSM using more than two overlapping images has some interesting characteristics. First of all, this approach strengthens the image orientation because of the redundancy in the geometric reconstruction. Points in object space can be calculated by the best fit of N convergent image rays instead of two. Secondly the redundancy leads to a more robust matching, as mismatches and a unique solution, in case of multiple matching candidates, can be easier identified. In the stereo case, an object point cannot be matched if it is located in an occluded area on one or both images. In the tri-stereoscopic case, the third image is taken from a different viewing angle. Consequently this leads to a shift of the occluded areas in the image and enlarges the chance of a successful match.

Processing of the Ikonos triplet is mainly done with a photogrammetric software platform, called SAT-PP. SAT-PP is able to perform image matching on more than two images simultaneously (Zhang & Gruen, 2006). This is in contrast to most photogrammetric software packages that are only able to match two images at the same time.

4.2 Preprocessing of the satellite data

Before processing the VHR imagery a contrast enhancement is executed as this leads to a more reliable image matching. Especially between images of the same area but taken at different dates from different orbits large radiometric dissimilarities can occur due to different illumination and atmospheric conditions, leading to poor matching results. To enhance the contrast for each image individually and to equalize the radiometric differences between the imagery, a Wallis filter was applied (Wallis, 1976).

The general form of a Wallis filter is given by:

$$g^w(x, y) = g(x, y) * r_1 + r_0 \qquad (1)$$

$$r_1 = \frac{cs_h}{cs_g + (s_h/c)} \qquad (2)$$

$$r_0 = bm_h + (1 - b - r_1)m_g \qquad (3)$$

with $g^w(x,y)$ and $g(x,y)$ = filtered and original image
m_g and s_g = original mean and standard deviation values
m_h and s_h = target value for mean and standard deviation
c and b   = contrast expansion and brightness forcing c^{te}

The Wallis filter performs a non linear, locally adaptive contrast enhancement. Actually a large kernel divides the image in different sections and within each section the local contrast is optimized. Applying a Wallis filter on the original images does not only result in an enhancement and sharpening of texture patterns in areas of low contrast and equal overall contrast but normalizes also the radiometry, especially between images taken at different dates. The effect of radiometric enhancement of very high resolution satellite imagery is illustrated in figure 1 & 2. The Wallis filter enhances existing texture patterns, leading to optimization of the contrast in shadow areas. Note that in the shadow rich areas axis-aligned artefacts are introduced due to the Wallis filtering.

Figure 1. Extract of original 11-bit Ikonos image, illustrating an area with high buildings. There is very little contrast within the shadow areas, leading to mismatches during the image matching process.

Figure 2. Extract of Wallis-filtered 11-bit Ikonos image. The radiometric filter enhances the existing texture patterns locally, leading to optimization of the contrast in the shadow areas.

Also an adaptive smoothing filter is applied to reduce image noise while sharpening edges. As noise is an important data-source for mismatches, reducing it further improves the quality of the surface model.

Next to the radiometric enhancement a method for geometric normalization was devised. The Ikonos 2002 stereo couple is epipolar projected and suitable for stereo applications. As the 2005 Ikonos image is taken from a different orbit, the images are displaced to each other and the internal geometry will be slightly different because of the different scan direction. Geometric normalization of the 2005 Ikonos image with the 2002 STEREO product imagery is done by image co-registration in ENVI. The 2005 image is resampled according a first-order polynomial transformation to geometrically align the multi-temporal imagery. A first-order polynomial transformation corrects for rotation, translation, scaling and shearing.

4.3 Bundle adjustment for image orientation

Rational polynomial coefficients are used to calculate initial values for internal and external image orientation. The rational polynomial function model uses a general polynomial transformation to describe the mathematical relationship between object and image space, instead of a physical sensor model. (Grodecki & Dial, 2003). As RPCs are calculated from on-board sensor orientation data, satellite ephemeris and star tracker observations, the accuracy of image orientation can be refined by using ground control points. During a field trip to Istanbul the necessary GCPs for photogrammetric processing of the DSM's were collected in close collaboration with the Istanbul Metropolitan Planning Centre (IMP-Bimtas). 37 clearly visible GCPs were collected, homogeneous distributed over the study area. In total, 17 points with known map coordinates and clearly identifiable in all three images were used to describe the relationship between the imagery and terrain. The a priori geometric accuracy for the DSM extraction consists of an overall RMSE value of 0.79 m for X residuals, 0.78 m for Y residuals and 2.36 m for Z residuals.

4.4 Epipolar geometry

Before extracting the surface model, the original images are resampled to an epipolar orientation. Y-parallax is removed, while leaving the parallax in X-direction unresolved, which can be interpreted as height differences. This reduces the process of finding conjugate points in overlapping images from a two-dimensional to a one-dimensional search algorithm along epipolar lines.

4.5 Multi-image matching

During the image matching process conjugate features need to be found automatically between the overlapping images. The surface model can be processed afterwards by calculation of height differences based on the measurement of the disparity between corresponding pixels. The applied algorithm works according a coarse-to-fine hierarchical matching strategy. Image pyramids consist of different versions of an image at exponentially decreasing resolutions. The bottom level of the pyramid contains the original image. The matching results of each higher pyramid level are used as approximations in the successive, lower level. At each level also an intermediate DSM is generated from the matched features and is refined through the image pyramid. Based on all data in each pyramid level, the matching parameters are fine-tuned progressively.

The matching algorithm is a combination of feature point, grid point and 3D edge matching. This redundancy leads to better constraints and more reliable results. Grid point matching is especially valuable in areas with less texture where conjugate feature points are hard to detect. For each grid point to be matched in the first image, the matching algorithm searches for the conjugate pixel in the other images that correlates the most by shifting a kernel of certain size along the epipolar line. A correlation constraint is used to identify possible matching candidates. The geometrically constrained cross-correlation or GC^3 method is an extension of the standard cross-correlation technique (Zhang & Gruen, 2006). In case of more than one matching candidate, the information of multiple images, i.e. more than two, can provide geometric constraints which assist to identify a unique matching solution.

3D edge matching is extremely valuable when dealing with urban areas, as they assist in modelling surface discontinuities. Edges are detected by the Canny operator (Canny, 1986). During surface model generation the matched edges will be taken into account as break lines to avoid smoothing effects. In Figure 3, illustrating matched edges in an urban area on Ikonos imagery can be seen that the main shape of most of the buildings is estimated quite well by detected edges. An important source of errors in edge detection is caused by building shadows. As shadow areas are being into large contrast with the surrounding pixels, edges will be detected at the shadow borders.

At a final stage a least-squares matching method, called modified multi-photo geometrically constrained matching algorithm, is performed using all matched points as approximations to detect mismatches and to further refine matching results. The MPGC algorithm combines the matched points with geometrical constraints, derived from multi-image ray intersection conditions and knowledge about the image orientation (Baltsavias, 1991). A Least Squares B-Spline Snakes is used to refine the matched edges. For more details on the matching strategy we can refer to (Zhang & Gruen, 2006).
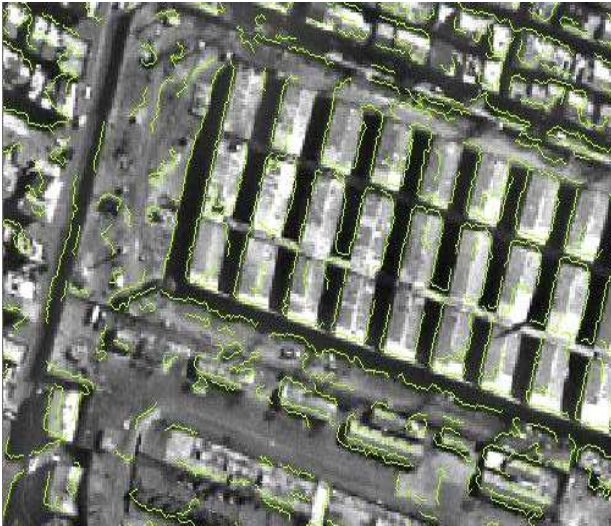
Figure 3. Edge detection & matching in an urban environment on Ikonos imagery.

During image matching, calculation of the position and height of each point or line is treated independently. To create a connected surface, the discrete measurements are interpolated. The resulting surface model is processed at a grid size of 3 meters. The chosen resolution leads to the best equilibrium between detail and reduction of noise. As illustrated in figure 4 & 5, the shape of big buildings and free-standing buildings is modelled well, while in the very dense urban area small buildings are merged into building blocks.
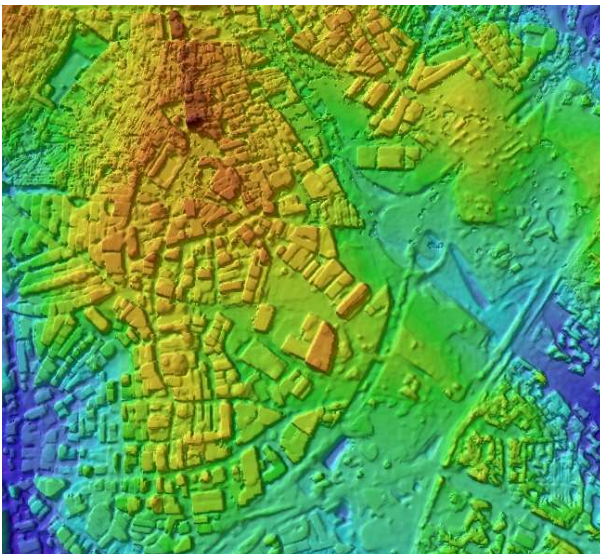

Figure 4. Map view on extract of the 3m colour-coded DSM.

4.6 Ortho-generation

During ortho-generation phase the sensor geometry of the images, characterized by a parallel projection in along-track direction and perspective projection in across-track direction, can be transformed to map geometry based on the developed surface model. The surface model represents each pixel in its correct geometric position. Back-projection from the DSM to the image supplies the grey value or texture for the pixel. In case of an occluded pixel on the master image, texture information is extracted from a slave image or neighbourhood

pixels in case of occlusion on all images. A ground sample distance of 1 m or 1 pixel is chosen for the ortho-image.
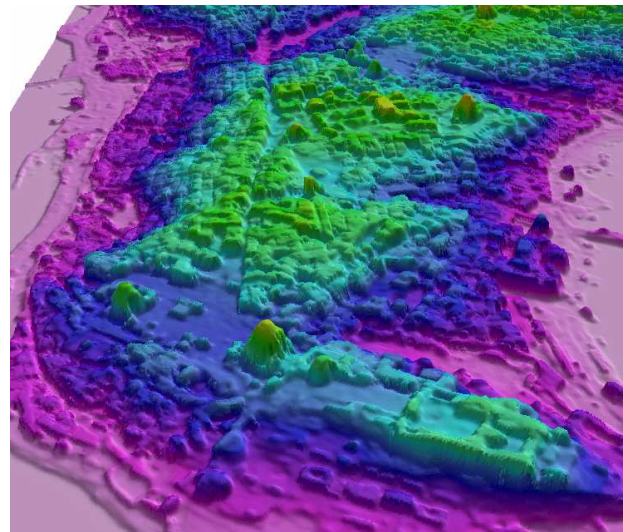

Figure 5. Perspective view on extract of 3m color-coded DSM. The surface model represents Istanbul's historic peninsula.

Figure 6. Extract of 3m surface model, draped with panchromatic ortho-image for photorealistic visualization.

5. SPATIAL FILTERING

To further improve the global quality of the surface model and especially to reduce smoothing effects, spatial filtering is applied on the height values of the DSM. A method is developed to further improve building shapes based on the knowledge of building contours. Flat roofs can be assumed for the buildings within the study field. An external dataset is used, consisting of 2D building footprints which were plotted on aerial imagery by IMP-Bimtas for cadastral purposes. Fitting of the 2D building footprints on the generated surface models, allows to extract all man-made objects. Subtraction of the DSM with the generated building model results in a terrain model (DTM) with gaps where the buildings were positioned. Distinction between a terrain model layer and a building model layer allows to apply different spatial filters adapted to the specific needs of the layer. The terrain model without man-made objects should be a continuous and smooth surface. As smoothing constraints are very important for the DTM, a median filter with a large kernel size of 18 by 18 pixels is used. On the other hand, smoothing must be minimized for the

building layer to model shape and discontinuities of man-made objects as good as possible. As the "bell-formed" shape of buildings in an unfiltered surface model is mainly an underestimation of height, an upper quartile filter with a small kernel of 7 by 7 is applied on the building layer two times within the boundaries of each footprint. An upper quartile filter is a nonlinear, order statistics filter and returns the 75th percentile value within the kernel. Spatial filtering of the height values within each building footprint reduces the local variations and puts the roof height on a more or less constant level. In a final step the DTM is merged with the building layer to obtain a final filtered DSM.

6. GEOMETRIC ACCURACY ANALYSIS

A dataset of 35 check points, measured with GPS and evenly distributed over the study area, is used to check the geometric accuracy and quality of the extracted models from stereopair and triplet. It concerns independent ground control points, meaning that they are not used in the photogrammetric processing of the models. Check points are preferred because of the lack of a more accurate reference surface model. Comparison of a measured height value and the calculated height value in the model at a certain location gives statistical information about the accuracy by which reality is modelled.
Some calculated statistics, quantifying the geometric accuracy are presented in table 2. Distinction is made between the standard stereoscopic and tri-stereoscopic approach. The a priori geometric accuracy reflects the quality and robustness of the image orientation. RMS error in X, Y and Z is given for the total of 17 ground control points that were used to fix the mathematical relationship between image and object coordinate space. For X and Y, sub-pixel accuracy is obtained in both approaches. RMSE for the Z component is less than 3 pixels. 35 independent check points are used to calculate the RMS error for Z and the mean Z difference between measured and calculated value by the model. For both statistics the value is less than 3 pixels.

	A priori geometric accuracy				DSM geometric accuracy		
Imagery	No. of GCP	RMSX (m)	RMSY (m)	RMSZ (m)	No. Of CP	RMSZ (m)	Mean dZ (m)
Stereoscopic	17	0.68	0.72	2.44	35	2.61	2.21
tri-stereoscopic	17	0.79	0.78	2.36	35	2.47	2.06

Table 2. Geometric accuracy analysis.

Visual analysis of the models shows big improvements of the quality for the surface model derived from the Ikonos triplet. Noise is reduced and smoothing effects of man-made object are reduced to a minimum, however the improvements do not reflect in the quantitative accuracy check. The RMSE and mean values are slightly better for the triplet than for the stereopair. This is due to the fact that the improvements are mainly situated around buildings and other steep changes in elevation. Check points are mostly measured in open terrain so that they are clearly identifiable on the imagery. Within these non-complex areas the surface model from the stereopair gives also optimal results. To have a better quantification of the improvements, future work should involve the collection of rooftop heights for a set of buildings and comparison between the collected ground truth and the produced models.

7. CONCLUSION

In this treatise an approach is proposed to extract an urban surface model in a semi-automatic way directly from multi-scopic Ikonos imagery, in contrast to surface models derived from manual plotting of building rooftops. As from the perspective of the geometric accuracy, as from the visual analysis we can conclude that the outcome is encouraging and that acceptable results are reached. At different levels of the photogrammetric processing of the imagery, efforts are done to cope with the complexity of modeling an urban environment. Occlusion and consequently mismatches are reduced by combining the redundant information of a third image with a stereopair. Radiometric and geometric dissimilarities between the multi-temporal imagery are diminished by preprocessing the individual images. Combination of three different matching algorithms gives redundancy and geometric constraints leading to dense and reliable matching results. Finally, spatial filtering is applied on the height values of the DSM to reduce smoothing effects and enhance global DSM quality.

REFERENCES

Baltsavias E., 1991. Multiphoto geometrically constrained matching. PhD Dissertation, Report No.49, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland.

Baltsavias E., Zhang L., Eisenbeiss H., 2006. DSM generation and interior orientation determination of ikonos images using a testfield in Switzerland. *Photogrammetrie, Fernerkundung, Geoinformation*, (1), pp. 41-54.

Canny J. F., 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8 (6), pp. 679-698.

Grodecki J., Dial G., 2003. Block Adjustment of High-Resolution Satellite Images Described by Rational Functions. *Photogrammetric Engineering and Remote Sensing* 69(1), pp. 59-70.

Jacobsen K., 2006. Digital surface models of city areas by very high resolution space imagery. *Workshop of the SIG Urban Remote Sensing*, Berlin, Germany, 2-3 March, 2006.

Raggam H., 2006. Surface mapping using image triplets : Case studies and benefit assessment in comparison to stereo image processing. *Photogrammetric engineering and remote sensing*, vol. 72, n° 5, pp. 551 – 563.

Wallis R., 1976. An approach to the space variant restoration and enhancement of images. *Proceedings of Symposium on Current Mathematical Problems in Image Science*, Naval Postgraduate School, Monterey, CA, 641-662.

Zhang C., Fraser C., 2008. Generation of digital surface model from high resolution satellite imagery. *Proceedings of XXIth ISPRS congress*, Beijing, China, 3 – 11 July 2008, pp.785-790.

Zhang L., Gruen A., 2006. Multi-image matching for DSM generation from Ikonos imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 60 (3) (2006), pp.195-211.