BUILDING HEIGHT ESTIMATION IN URBAN AREAS FROM VERY HIGH RESOLUTION SATELLITE STEREO IMAGES

A. Alobeid, K. Jacobsen, C. Heipke

Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover
Nienburger Str. 1, D-30167 Hannover, Germany
[Alobeid, Jacobsen, Heipke]@ipi.uni-hannover.de

KEY WORDS: satellite data, digital surface model, urban area, local and global matching, empirical comparison, height estimation

ABSTRACT:

The extraction of the third dimension from stereoscopic image pairs is a well known technique. Since in a number of countries aerial images and laser scanner data are unavailable, expensive or classified, high resolution optical satellite images provide a viable alternative to generate digital surface and digital terrain models. Especially the automatic extraction of highly accurate 3D surface models in urban areas is still a very complicated task due to occlusions, large differences in height and the variety of objects and surface types.
In this paper we present an analysis and a comparison of three different matching methods for generating urban DSMs based on very high resolution satellite images: least squares matching (LSM; Förstner 1982) in a region growing fashion, dynamic programming (DP) according to Birchfield, Tomasi (1999), and semiglobal matching (SGM; Hirschmüller, 2008). We empirically study the effects of the three methods as applied to three different IKONOS stereo pairs with a ground sampling distance of 1.0m.
It comes as no surprise and can be clearly seen in our results that in the LSM result the shape of the buildings is considerably smoothed. While in the DP results the building shape is sharper, only little detail is visible on the roof. With SGM more details are visible, but the result seems to contain some artefacts. As far as geometric accuracy is concerned we found based on independent manual checks, that for all three methods the height accuracy is in the range of the ground resolution of 1.0m, corresponding to 0.6 pixels x-parallax given a h/b ratio of 1.7. This value includes not only the matching accuracy, but also the accuracy of manual measurement. Thus, the accuracy of the automatic matching is better than 1.0m in height.

1. INTRODUCTION

Digital surface models (DSMs) of urban areas are becoming increasingly important for many applications, e. g. tele-communication (Renouard et al., 1999), urban planning (Kux et al., 2006), map updating (Caetano, 2001) and monitoring urban growth which occurs very fast especially in second or third world cities. Stereo pairs from very high resolution satellites such as IKONOS, QuickBird, WorldView and GeoEye-1 led the way into a new era of generating urban DSMs, not limited by restrictions for the access to aerial images and laser data.
Manual data acquisition is often too time consuming and thus too expensive. As a consequence an automatic procedure for the generation of DSMs including building shapes based on image matching techniques is highly desirable. Several methods were suggested in the literature and have given satisfactory results based on high resolution satellite imagery (e. g. Krauss et al., 2008, 2005; Büyüksalih, Jacobsen, 2007; Poon et al., 2007; Jacobsen, 2006; Zhang, Grün, 2006, 2004). Nevertheless, the extraction of accurate urban DSMs is still an unsolved problem, partly due to occluded areas, sudden changes in height and the large variety of objects and surface types. As a consequence in particular near building boundaries difficulties and limitations still exist.
In this paper we present an analysis and a comparison of three different matching methods for generating urban DSMs based on very high resolution satellite images:

- least squares matching (LSM; Förstner, 1982) in a region growing fashion (Otto, Chau, 1989; Heipke et al., 1996), a local area based method which compares the intensity values within a template to those in a search window;
- dynamic programming (DP) according to Birchfield, Tomasi (1999), a global method which determines pixel disparities in epipolar lines by searching for a best path through the related cost matrix based on individual pixel intensity values as input for a dissimilarity measure;
- semiglobal matching (SGM; Hirschmüller, 2008), which computes conjugate points along multiple conjugate lines hierarchically by using mutual information instead of intensity value differences as dissimilarity measure.

Section 2 of this paper presents the characteristics of the input data used for our study. The three matching algorithms are shortly reviewed and results are presented in section 3. Section 4 is dedicated to an accuracy analysis of the three matching methods. Finally, the paper is concluded with section 5.

2. STUDY AREA AND USED TEST DATA

Three study areas were examined, located in the cities of Maras and Istanbul in Turkey and San Diego, USA. The study areas are located in rolling terrain, containing densely built up parts and some single buildings with heights up to 65m.
For each area a panchromatic IKONOS GEO stereo model with a ground sampling distance (GSD) of 1m was available. The height-to-base (h/b) ratio for Maras is about 7.5 (angle of convergence 7.5°), reducing occlusions, while the h/b value is 1.6 for Istanbul (angle of convergence 35°) and 1.7 for San Diego (angle of convergence 32°), enlarging the disparities and causing some additional matching problems due to the different perspective (see Figure 1). The sun elevation is 50.8° for Maras, 65.5° for Istanbul and 34.2° for San Diego.

From the IKONOS stereo pairs quasi-epipolar images have been generated as required for DP and SGM.
A number of sub-areas with different characteristics have been selected based on building shapes and density. Identical sub-areas were used for the investigations of the three matching methods. Forest areas and water surfaces were not included in the sub-areas.

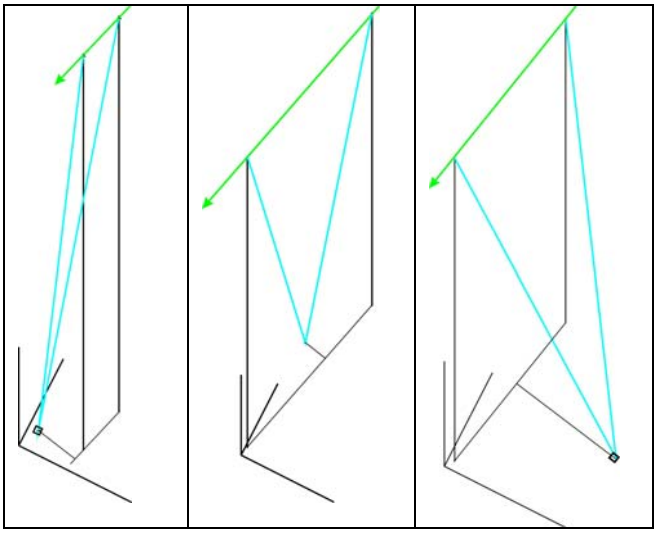

Figure 1: imaging geometry for the three test sites; from left to right: Maras, Istanbul, San Diego

3. DESCRIPTION OF MATCHING METHODS

3.1 Least squares matching with region growing

This local area based method (Förstner, 1982) uses the intensity values to estimate the disparity for the centre pixel of a template window. An affine transformation allows for geometric distortions of the template before comparing the intensity values with those of the second image. Based on the principle of least squares the sum of the squared differences of the corresponding normalized intensity values is then minimized, the affine parameters serve as unknowns; see eq. (1) and (2).

$$v(x,y) = g(x,y) - [r_o + r_1 \cdot g'(x',y')] \quad (1)$$

$$x' = a_o + a_1 x + a_2 y$$
$$y' = b_o + b_1 x + b_2 y \quad (2)$$

where:

x,y:	coordinates of template window
x´,y´:	coordinates of second image
g(x,y):	intensity values of template window
g´(x´,y´):	intensity values of the second image
r_0, r_1:	parameters for intensity value normalisation
a_0,a_1,a_2,b_0,b_1,b_2:	affine transformation parameters
v(x,y):	residuals of normalized intensity value differences

The region growing strategy, published by (Otto, Chau, 1989; Heipke et al., 1996) requires a few seed points, which in our case were provided manually. The algorithm starts from a seed point, and then matches the four neighbouring points at a pre-defined step size, subsequently continuing with the point with highest correlation coefficient. To be accepted, the correlation coefficient had to be larger than 0.8 for Maras and San Diego, and larger than 0.6 for Istanbul. These thresholds are based on intensive tests (Jacobsen, 2006; Alobeid, Jacobsen, 2008).
Due to occlusions, the region growing strategy may stop in areas that contain large height differences (large x-parallaxes) caused e. g. by buildings. Also, due to the use of a window of constant shape (a plane in 3D space), the algorithm is not able to track building outlines, leading to smoothed building forms because some pixels in the template may be located on top of the building and others on the ground or a wall. In particular for larger window sizes, adjacent buildings may also be merged and appear as a common blob.
We found empirically that the optimal size of the matching window in our test areas was 10 by 10 pixels. While smaller windows are influenced by noise, although they may give a more precise building shape, larger windows smooth the DSM significantly. In the case of a small angle of convergence, as is the case in Maras with only 7.5° convergence angle, the images of a stereo scene are rather similar, resulting in very high correlation coefficients and a good success rate of matching.
Figures 2, 3 and 4 show digital surface models calculated by least squares matching with region growing for Maras, Istanbul and San Diego.
For Maras between 89% and 94% of all possible points could be successfully matched. Moreover, between 86% and 93% of the points had a correlation coefficient exceeding 0.95. Due to the low image quality and a smaller h/b ratio of 1.6 (35° angle of convergence) in Istanbul, between 56% and 76% of all possible points could be successfully matched. Between 62% and 71% of the points had a correlation coefficient exceeding 0.60. The better image quality for San Diego led to a success rate of between 72% and 83%. Between 65% and 78% of the points had a correlation coefficient exceeding 0.8.
Obviously, these values must be seen relative to the number and distribution of manually provided seed points.

Figure 2: DSM generated by LSM, Maras test site (window size 540x465 pixels)

When visually inspecting the obtained results, the impression is that least square matching usually provides a dense and accurate disparity map with only few blunders. However, the fixed template shape is not able to faithfully extract building outlines. Instead, the buildings indeed appear as low path filtered blobs, as discussed above.

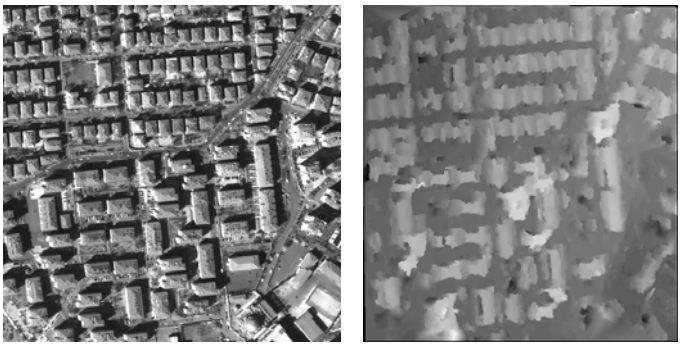

Figure 3: DSM generated by LSM, Istanbul test site (window size 500x500 pixels)

Figure 4: DSM generated by LSM, San Diego test site (window size 600x700 pixels)

3.2 Dynamic programming

The reported drawbacks in local area based methods prompted us to think about an alternative solution. As the first possibility a matching algorithm for epipolar images based on dynamic programming (Birchfield, Tomasi, 1999), has been chosen. No windows are required for matching, as intensity values of individual pixels are compared, combined with constraints to reward successful matches and to penalise occlusions. The algorithm focuses specifically on generating correct results at height discontinuities, sacrificing some accuracy in smooth areas.

The matching problem is expressed as an optimisation problem for each corresponding epipolar line pair; based on a pre-defined cost function. Each pixel in the left epipolar line is compared to all pixels of the conjugate epipolar line, and a 2D array of costs is constructed. The used cost function $\lambda(x,y)$ has three components, see eq. (3):

$$\lambda(x,y)= \sum_{i=1}^{N_m} d(x_i, y_i) - N_m * K_r + N_{occ} * K_{occ} \qquad (3)$$

- The first component is a sum of the dissimilarities $d(x_i,y_i)$ between the matched pixels, it should dominate the cost function.
- The second component (N_m*K_r) is a reward for correct matching, where N_m is the number of matched pixels and K_r is the match reward per pixel.
- The third component ($N_{occ}*K_{occ}$) is a penalty for occlusions, where N_{occ} is the number of occlusions and K_{occ} is the occlusion penalty.

The easiest dissimilarity function is the absolute value of difference in intensities. Instead, we compute the dissimilarities by using linearly interpolated intensities halfway between each pixel in each corresponding epipolar line and its neighbours according to Birchfield and Tomasi (1998a) to reach sub-pixel accuracy over all possible disparities.

The algorithm then computes the sequence of all best corresponding pixel pairs according to minimal cost. For each path through the 2D array of costs, the total cost is calculated according to eq. (3). Then, the optimal path with minimal cost is determined using dynamic programming.

Again, the parameters of the method have been determined empirically. We first impose a threshold for the maximum disparity; this value should exceed the maximum disparity in the scene.

K_r and K_{occ} are the remaining values to be selected. K_r is the maximum amount of pixel dissimilarity expected between two correctly matching pixels. K_r has been varied in the range [3 - 12] with less then 3% of all pixels changing their disparity value. K_{occ} is the evidence to declare an occlusion and thus a change in disparity. We varied K_{occ} in the range [13 - 46], in these tests less then 9% of all pixels changed in disparity.

In our study, we found empirically that optimal values for K_r and K_{occ} were 7 and 12 in Maras and 5 and 20 in Istanbul, 7 and 35 in San Diego respectively based on visual inspection.

The results for the test sites are shown in figures 5, 6 and 7. The results show the algorithm's ability to provide a dense coverage of corresponding pixels and thus to compute an approximate disparity image, particularly at building outlines. As is generally known for matching epipolar lines independently and as can be clearly seen in the left part of figures 5, 6, and 7, a streaking effect appears in the epipolar direction, causing distortion of building borders.

The results have been post-processed by median filtering in the vertical direction. With a 7x1 window the best improvement could be reached as judged from visual inspection. Based on this vertical median filter, the shape of buildings becomes clearer, see right part of figures 5, 6 and 7.

In densely built-up areas, such as Istanbul, the buildings close to each other appear as building blocks where the hidden parts are not matched

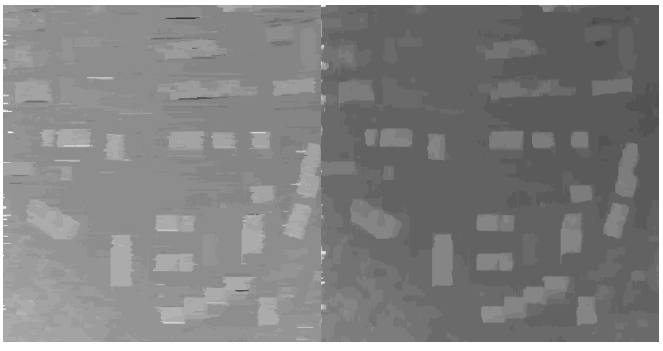

Figure 5: DSM generated by DP, Maras test site
Left: Result after epipolar matching
Right: Result after vertical median filtering

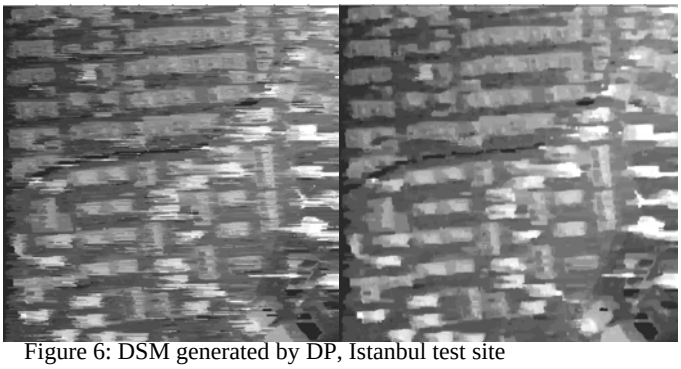

Figure 6: DSM generated by DP, Istanbul test site
Left: Result after epipolar matching
Right: Result after vertical median filtering

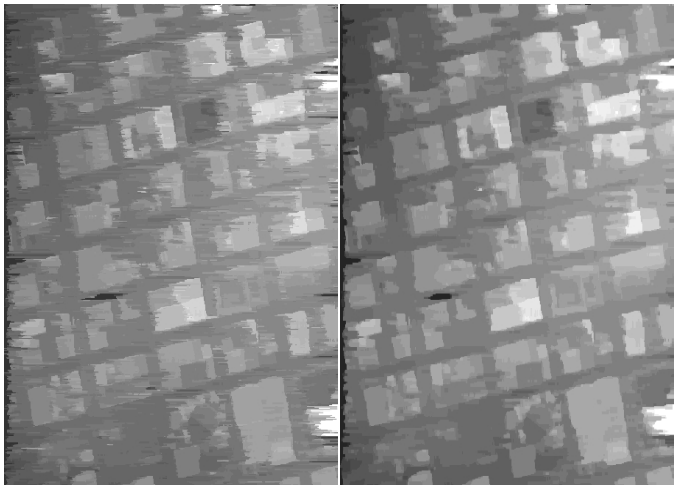

Figure 7: DSM generated by DP, San Diego test site
Left: Result after epipolar matching
Right: Result after vertical median filtering

3.3 Semiglobal matching (SGM):

Semiglobal matching (SGM; Hirschmüller, 2008) is an extension of the previous method and incorporates a smoothness constraint that is usually expressed as a global cost function, for determining the disparities of several line pairs intersecting in one pixel simultaneously by applying energy minimization techniques.

It is based on the two following ideas: First, the dissimilarity is expressed pixel per pixel by Mutual Information (MI). MI measures correspondence without assuming that conjugate points have similar intensity values. Instead, the joint probability distribution in the form of the joint intensity value histogram is used. MI has been shown to be rather robust with respect to radiometric differences. Good descriptions of MI can be found in Kim et al. (2003) and Egnal (2000).

Second, a global 2D smoothness constraint across multiple intersecting lines is introduced. It is approximated by combining many 1D constraints.

The first step for SGM is to obtain an initial disparity image that is required for warping one of the stereo images before MI can be calculated. In line with Hirschmüller (2008) we start with a random disparity image, and then continue in a hierarchical fashion.

Subsequently, the joint histogram is derived over the whole images. It is stored as a 256×256 histogram and is smoothed using a Gaussian kernel. Then, the MI values are computed.

The third step is to determine a disparity image that minimizes the energy function by pathwise optimization of several 1D-paths toward the pixel under consideration. Thereafter, the costs are summed over all paths ending in this pixel, see eq. 4:

$$S(p,d) = \sum_{i=1}^{r} Lr(p,d) \qquad (4)$$

$$Lr(p,d) = C(p,d) + \min[Lr(p-r,d), Lr(p-r,d-1) + P1,$$

$$Lr(p-r,d+1) + P1, \min_i Lr(p-r,i) + P2] - \min_k Lr(p-r,k)$$

where:

p:	image location of current pixel
d:	disparity value $d \in [d_{min}, d_{max}]$
S (p,d):	aggregated cost
C (p,d):	pixelwise matching cost
Lr (p,d):	cost paths toward the actual pixel
P1:	a small value penalising disparity changes between neighbouring pixels of one pixel.
P2:	a large value penalising disparity changes of more than one pixel between neighbouring pixels.
r:	number of accumulated paths (according to Hirschmüller, r should be 8 or 16).

The first component is the pixel-wise matching cost from MI, while the remaining components in the equation add the lowest cost of the previous pixel of the path. In our study, we found empirically, again based on visual inspection, that optimal values for P1 and P2 were 4 and 8 in Maras and 5 and 9 in Istanbul, 6 and 11 in San Diego.

The final disparity image is then computed according to eq. (5):

$$D(x,y) = \min_d \left\{ \sum_{i=1}^{r} Lr(x,y,d) \right\} \qquad (5)$$

The SGM results with the same three test sites as for the other three methods are shown in Figures 8, 9 and 10. It can be seen very clearly that indeed streaking is much reduced as compared to the dynamic programming results.

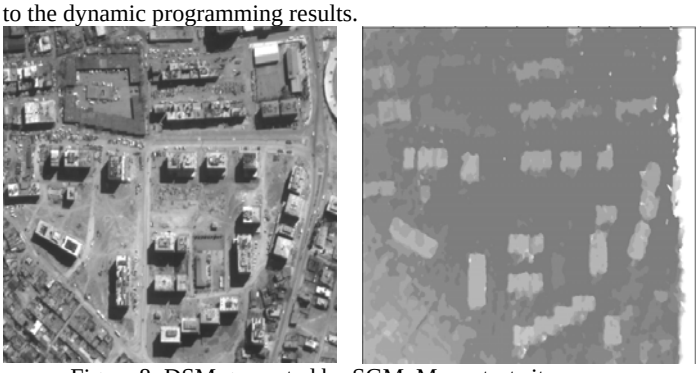

Figure 8: DSM generated by SGM, Maras test site

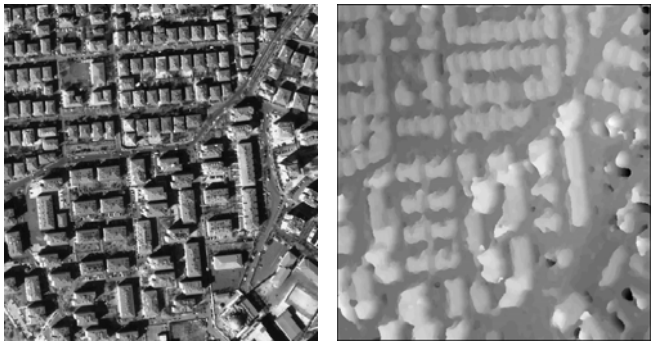

Figure 9: DSM generated by SGM, Istanbul test site

Figure 10: DSM generated by SGM, San Diego test site

Figure 11 shows a 3D view of part of the results of the San Diego test site. It can be seen that in LSM no clear building shape were generated and occlusions partly cause tilted facades. The DP result has to be post-processed, otherwise there is too much streaking. SGM reveals more detail than DP such as roof structure. The reason is that in SGM optimization is done for every pixel, whereas in DP the epipolar lines or at least parts of the epipolar lines are assigned constant disparity. As a consequence, SGM is able to reconstruct e. g. gable roofs, whereas DP cannot.

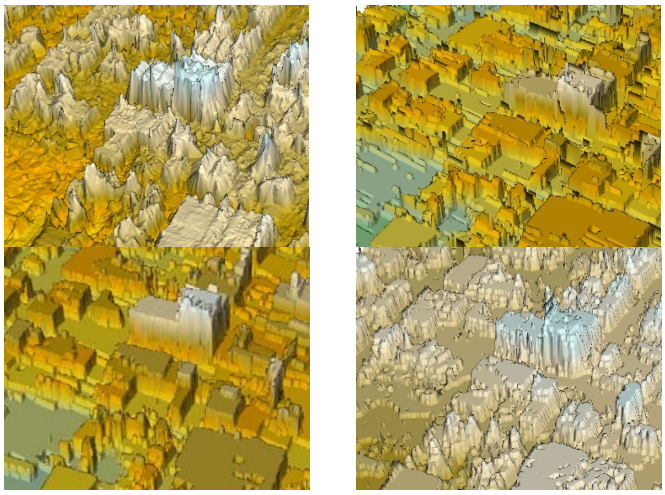

Figure 11: 3D-view to DSMs, San Diego test site
Upper left: LSM Upper right: DP without filtering
Lower left: DP filtered Lower right: SGM

To study the behaviour of the three matching methods in more detail, we present the results of one building of the San Diego test site in an enlarged view, see Figure 12.
It is very clear that in LSM result the outline of the building is considerably smoothed. While in the DP results the outline is sharper, only little detail is visible on the roof of the building. With SGM more details are visible, but the result seems to contain some artefacts.

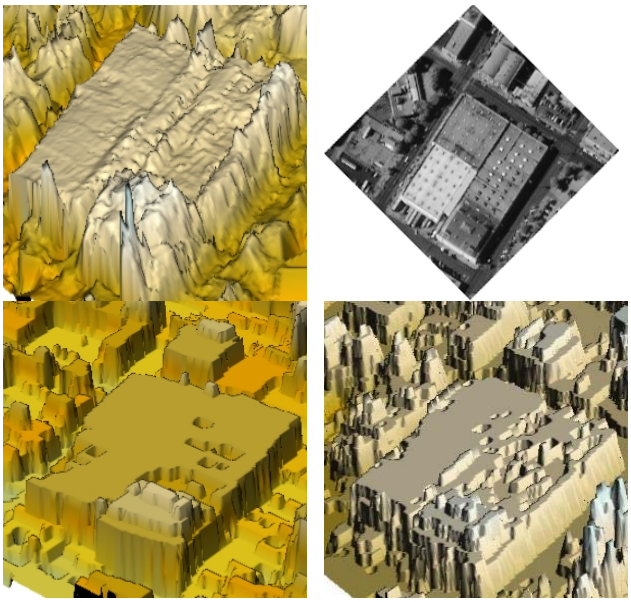

Figure 12: 3D-view to one building, San Diego test site
Upper left: LSM Upper right: real building
Lower left: DP filtered Lower right: SGM

4. ACCURACY ANALYSIS

In order to independently check the geometric accuracy of the obtained results, building heights and points on the ground have been measured manually in the San Diego stereo model. These measurements have been compared with the generated height models. Especially for the DSM from least squares matching, the discrepancies against the reference data depend on the point location in relation to the facades due to the discussed smoothing effect. To minimise these problems we selected points in the centre of the building tops and on the ground with a sufficient distance from the facades.
The root mean square (RMS) difference between the manually measured heights and the LSM results turned out to be +/- 1.0m, for DP the RMS difference was +/- 1.2m and for SGM +/- 0.7m. The variation of the RMS values are probably not significant, it depends on location of single points having larger discrepancies. A direct comparison of the matched data showed only negligible differences below 0.2 pixels in the x-parallax on flat parts as open ground and flat roof tops. Differences between the methods are mostly visible at the facades.
In general it can thus be stated, that for all three methods the height accuracy is in the range of one pixel GSD or 1.0m, corresponding to 0.6 pixels x-parallax given the h/b ratio of 1.7. This value includes not only the matching accuracy, but also the accuracy of manual measurement. Thus, the accuracy of the automatic matching is better than 1.0m in height.

5. CONCLUSION

The generation of digital surface models in urban areas based on IKONOS stereo pairs has reached a high level of accuracy. The three investigated methods show differences in detail and shape, the overall geometric accuracy is rather similar.
The area based least squares matching is not able to generate clear building outlines and strongly depends on occlusions. Dynamic programming requires post-processing across the

epipolar lines to reduce streaking while this is not required for semiglobal matching.
These findings should be seen as a first result of our study on comparing different matching methods applied to urban areas. In future work we will investigate larger test sites, look at the occurrence and the elimination of blunders and also take performance issues into account.
Furthermore, the impact of the different dissimilarity measures will be studied. In order to combine the advantages of both methods we plan to investigate the SGM approach based on the dissimilarity measure suggested by Birchfield and Tomasi and used in our DP experiments.

REFERNCES

Alobeid, A.; Jacobsen, K., 2008: Automatic generation of digital surface models from Ikonos stereo imagery and related application: GORS, 16th International Symposium. Damascus, Syria, on CD: also available at www.ipi.uni-hannover.de (last access April 2009)

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

Birchfield, S., Tomasi, C., 1998b: A pixel dissimilarity measure that is insensitive to image sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(4):401-406

Birchfield, S., Tomasi, C., 1999: Depth discontinuities by pixel-to pixel stereo. International Journal of Computer Vision 35(3): 269-293

Büyüksalih, G.; Jacobsen, K., 2007: Digital Surface Models in Build up Areas based on very high resolution Space Images: ASPRS annual conference. Tampa, on CD, also available at www.ipi.uni-hannover.de (last access April 2009)

Caetano, M., Santos, T., 2001: Updating land cover maps with satellite images, Geoscience and Remote Sensing Symposium, IGARSS '01. IEEE**,** vol.3, pp. 979-981

Egnal, G., 2000: Mutual Information as a Stereo Correspondence Measure, Tech. Rep. MS-CIS-00-20, University of Pennsylvania

Förstner, W., 1982: On the geometric precision of digital correlation, IntArchPhRS, (24)3, pp. 176- 189

Heipke, C., Kornus, W. and Pfannenstein, A., 1996: The evaluation of MEOSS airborne 3line scanner imagery processing chain and results. Photogrammetric Engineering & Remote Sensing 62(3):293–299

Hirschmüller, H., 2005: Accurate and Efficient Stereo Processing by Semiglobal Matching and Mutual Information, IEEE Conf. on Computer Vision and Pattern Recognition CVPR'05, Vol. 2, San Diego, CA, USA, pp. 807-814

Hirschmüller, H., 2006: Stereo Vision in Structured Environments by Consistent Semiglobal Matching, IEEE Conf. on Computer Vision and Pattern Recognition CVPR'06, Vol. 2, New York, NY, USA, pp. 2386-2393.

Hirschmüller, H., 2008: Stereo Processing by Semiglobal Matching and Mutual Information, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2):328-341

Jacobsen, K., 2006: Digital surface models of city areas by very high resolution space imagery, EARSeL Workshop on Urban Remote Sensing. Berlin, on CD, also available at www.ipi.uni-hannover.de (last access April 2009)

Kim, J., et al., 2003: Visual Correspondence Using Energy Minimization and Mutual Information, IEEE Int. Conf. Computer Vision, 2003, Vol. 2, pp. 1033- 1040

Krauß, T., Lehner, M., Reinartz, P., 2008: Generation of coarse 3D models of urban areas from high resolution stereo satellite images, IntArchPhRS. Vol. XXXVII. Part B1, pp. 1091-1098

Krauß, T., et al., 2005: DEM generation from very high resolution stereo satellite data in urban areas using dynamic programming, IntArchPhRS, Vol. 36 (1/W3), on CD, ISPRS Hannover Workshop

Kux H., et al., 2006: High-Resolution Satellite Images for Urban Planning, studies in Progress at INPE (National Institute for Space Research), Brazil, IntArchPhRS, Vol. XXXVI - Part 2, pp.121-124

Otto, G. P., Chau, T. K. W., 1989: Region-growing algorithm for matching of terrain images. Image and Vision, 7(2): 83-94.

Poon, J., Fraser, C., Zhang, C., 2007: Digital surface models from high resolution satellite imagery. Photogrammetric Engineering & Remote Sensing, 73(11):1225-1232

Renouard, L., Lehmann, F., 1999: High resolution Digital surface Models and Orthoimages for Telecom Network Planning. Fritsch D. & Spiller R. (eds.), Photogrammetric Week`99`, Wichmann, Heidelberg, pp.241-246

Zhang, L., Grün, A., 2004: Automatic DSM generation from linear array imagery data. IntArchPhRS, 35(B3): 128-133

Zhang, L., Grün, A., 2006: Multi-Image Matching for DSM Generation from IKONOS Imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 60(3):195-211