REFINEMENT OF URBAN DIGITAL ELEVATION MODELS FROM VERY HIGH RESOLUTION STEREO SATELLITE IMAGES

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

Commission I, WG 4

KEY WORDS: VHR satellite data, stereo images, digital surface models, optimization, dense stereo matching, occlusion mask, blunder elimination

ABSTRACT:

Digital elevation models (DEM) of high resolution and high quality are required for many applications like urban modeling, readiness for catastrophes or disaster assessment. A good source for the derivation of such DEMs from any place in the world are very high resolution (VHR) satellite stereo images as provided e.g. by Ikonos, QuickBird or WorldView. In this paper a method for the generation and refinement of urban high resolution DEMs from VHR imagery is presented and evaluated. Urban DEMs generated from very high resolution satellite imagery of ground sampling distances of about one meter are normally of resolutions of about three to ten meters. For the above mentioned applications of urban DEMs such results are often too coarse. In this paper an advanced method for the generation of dense digital elevation models is presented and discussed. The method is mainly based on dense stereo algorithms developed for computer vision applications. It is adapted and optimized to earth observation requirements. In the paper the DEM generation together with the additional refinement steps is presented and evaluated using very high resolution stereo imagery from Munich and Athens. The generated DEMs are compared to ground truth data where available and the quality and efficiency of the algorithms are analyzed and discussed.

1 INTRODUCTION

High quality digital elevation models (DEM) provide essential basic information for many tasks. Especially in urban areas such DEMs provide basic data for the generation of urban models. DEMs occur as digital surface models (DSM) describing the visible surface, as digital terrain models (DTM) representing the ground or as normalized DEMs (nDEM) containing the elevation of objects above the DTM. The generation of digital surface models from very high resolution optical stereo satellite imagery delivers either rather good DSM of relatively coarse resolution of about 1/3 to 1/10 of the original ground sampling distance [Lehner et. al., 2008] or high resolution DSM with many mismatches and blunders [Xu et. al., 2008].

To overcome this dilemma many authors suggest on one hand methods for more stringent image correlation approaches in the DSM creation or on the other hand methods for blunder detection and removal directly from the DSM afterwards.

In the presented paper we will follow a combined and object oriented approach fusing the generated DSM with the original imagery.

Classically so called automatic terrain extraction (ATE) methods as implemented in many software packages like e.g. ERDAS [Xu et. al., 2008] are used for the generation of DSMs from aerial or satellite imagery. These methods are already mature and deliver results of high accuracy for input data consisting of two or more images. The classical ATE extracts in a first step so called interest points [Förstner and Gülch, 1987] in one of the images and locate these using an area based pyramid matching within the other images [Lehner and Gill, 1992]. After densifying the point cloud using region growing [Otto and Chau, 1989] a least square matching produces a huge amount of high quality image to image correlation points. Applying forward intersection using the camera parameters or provided RPC creates 3D points which can be triangulated to a regular grid – the resulting DSM.

Besides this widespread method in remote sensing other methods e.g. from computer vision show interesting alternatives. Many of these methods are subject to the taxonomy and evaluation done by [Scharstein and Szeliski, 2002] and are mainly dense stereo algorithms performing best on small epipolar images. Following the investigations done by [Pentenrieder, 2008] we selected for our approach the two best ranked methods – namely the digital line warping [Krauß et. al., 2005] and the semiglobal matching [Hirschmüller, 2005]. Fusing these algorithms (more stringent image correlation) and adding additional object orientated blunder detection leads to the presented method.

2 METHOD

2.1 DSM generation

The generation of the digital surface model (DSM) followed in this paper is based on an hybrid approach fusing the digital line warping as described in [Krauß et. al., 2005] and the semi global matching described in [Hirschmüller, 2005]. Both methods are actually not yet very common in remote sensing. Their origin lies in speech recognition and computer vision. Both algorithms work only on one stereo image pair (two images, no multiple image stereo) in epipolar geometry.

But on the other hand both algorithms provide a dense stereo matching. This is in contrast to the usual ATE methods in remote sensing which perform a matching of only very good interest points between two or more images followed by a triangulation to deliver the regular DSM.

Both algorithms are based on a (global) optimization of matching costs. The calculation of the matching costs can be done by different approaches. The simplest approach is the absolute difference of the gray values of the images yielding much better results than e.g. the squared difference. Another approach is the Birchfield-Tomasi distance [Birchfield and Tomasi, 1998] including the minimum of the absolute difference of the first pixel to the second pixel and within half the distance to it's neighbors.

Another approach is the usage of probability values instead of the gray values of the two pixels – the so called mutual information. These values are based on entropy calculations between the correlated images. A good description can be found in [Hirschmüller, 2008].
Beside the pixel by pixel calculation of the cost also the usage of small areas around every pixel is possible independent of the cost function used. For all further investigations only a pixel by pixel Birchfield-Tomasi correlation not using any larger areas is used if not mentioned otherwise.
In a dynamic programming approach the costs are summarized for all possible disparities and the disparity with the lowest summarized cost is selected. The digital line warping correlates images line by line and delivers a disparity vector warping one line to the other. Due to this line by line approach the generated DSMs suffer from streaking effects due to the missing linkage between scanlines which is shown in Figure 1, left.
The semiglobal matching approach introduced by [Hirschmüller, 2005] overcomes this flaw by summarizing costs not only in the direction of the epipolar line but in 8 or 16 directions across the full image. Since the dynamic programming approach only works for a two dimensional matrix (position in epipolar line vs. disparity) the cost aggregation is done for each of the 8 or 16 directions separately and the aggregated costs are summarized afterwards..

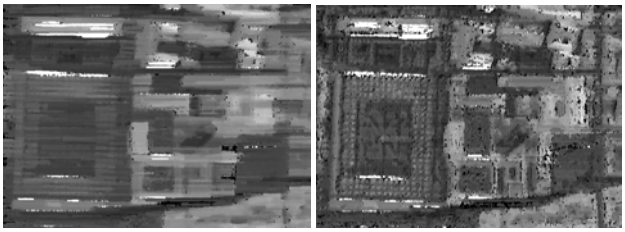

Figure 1. DSM generated using no layers (left) and two layers (right)

But this approach is very time consuming since also costs produced by scanlines far away from the actual line need to be considered. Due to this a hybrid approach using only the neighboring scanlines – so called layers – is realized. In this paper (unless noted otherwise) a standard two-layer-configuration is used with two scanlines before and two scanlines after the considered epipolar line accounting to the summarization of costs. This is sufficient for eliminating the streaking effects of the digital line warping process as shown in Figure 1, right.
An additional optimization to the fusion approach is the usage of an existing "coarse" disparity map. Such coarse disparity maps can be generated by using image pyramids. If such a previous disparity map is given their values are taken into account as offsets in the calculation of the actual disparity. In this approach really large disparity ranges of about ±400 px which represent nearly 2000 m height difference in standard Ikonos stereo configuration can be scanned very fast using only small disparity ranges if the processing starts in a high image pyramid level.

2.2 Blunder elimination

The described fused DSM generation algorithm also delivers information on the quality of derived disparities. So in a forward-backward check inconsistent disparities can be detected and eliminated.
This is done directly after the aggregation of matching costs via the dynamic programming approach. The disparities *d(x)* are calculated relative to the pixel positions *x* in the left image of the stereo pair. Projecting every disparity value of the actual epipolar line to it's counterpart position of the right stereo image results in a value $-d(x)$ at position $x+d(x)$ of the right image. This is illustrated in Figure 2. The disparity values of the left image are shown at the left (blue) pixels. The number in each pixel in the diagram represents the disparity *d(x)* calculated for this position. The correlated epipolar line of the right image is shown on the right (green) with *d(x)* at positions $x+d(x)$. If there exist blunders a two or more disparities $d(x_1)$ and $d(x_2)$ refer to the same pixel in the right image: $x_1+d(x_1)=x_2+d(x_2)$. Filtering out such collisions and reprojecting the disparities from the right image back to the left image results in a blunder filtered disparity image. The blunders detected by this approach are mostly occlusions – image parts which can only be seen in one image of the stereo pair.

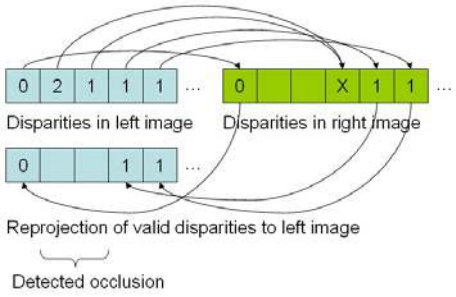

Figure 2. Occlusion detection by forward-backward check during DSM generation

In a second check the matching of the left to the right image is reversed (the right image becomes the base image for the disparity calculation) and the generated reverse disparity image gets also reprojected to the previous disparity image excluding additionally disparities which differ more than a given threshold (one pixel for default).
Since the generated disparity image fits on the left stereo image such eliminated disparities depict areas which can be seen in the left image (per definitionem) but are occluded in the right image.

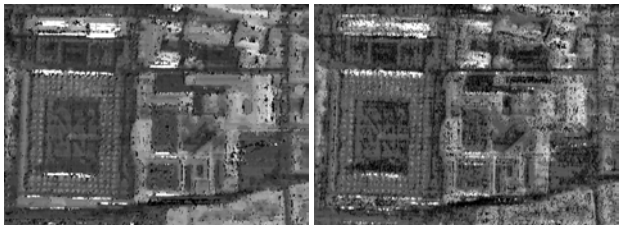

Figure 3. Left: left to right DSM with occluded regions from the right imaged marked as voids, right: right to left DSM with occluded regions from the left image marked as voids (reprojected to left; both Munich, Residenz)

Doing an orthorectification of the left stereo image to a geographic coordinate system shifts the disparities to correct positions and heights using the geometry from the satellite imagery provided as rational polynomial coefficients (RPCs). This shift introduces a second kind of occluded pixels. These are pixels which are not visible in the left image (see Figure 4).
Both types of occlusions in the ortho rectified DSM get filled in the ortho rectification process from the lowest neighbouring disparity in epipolar direction but remain remembered in an additionally generated occlusion mask. This mask is used afterward for the creation of a true ortho image. Occlusions

originating from the "fw/bw" test described above can be filled in the ortho image from the left stereo image whereas occlusions generated in the DSM ortho rectification ("ortho") can be filled from the right stereo image.

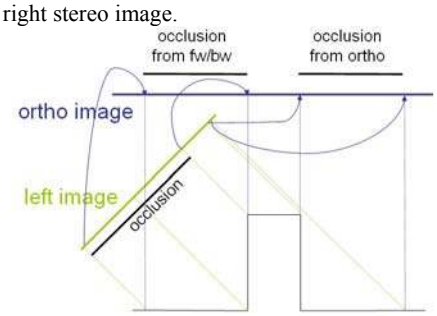

Figure 4. Origin of the two distinguishable types of occlusions in the resulting ortho image

Afterwards a DSM fill process using an adaptive median filter is done. In this process only void pixels (occlusions) get filled. So all blunders remain in the filled image. A subsequent median filtering of the filled image and subtracting this result from the filled image give the position of possible blunders. Applying a threshold can eliminate such small blunders. Repeating the adaptive median fill procedure with the blunder corrected original DSM give a rather good surface model (Figure 5).

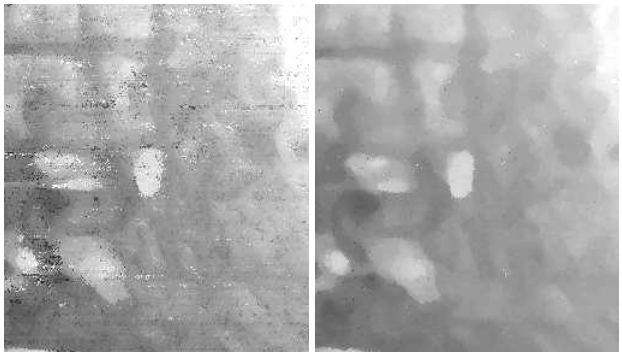

Figure 5. Blunder elimination by adaptive median fill; left: filled DSM with blunders, right: filled DSM after eliminating blunders (Athens, Acropolis)

2.3 DTM extraction

There exist many approaches for the extraction of digital terrain models representing the bare earth from digital surface models, mostly developed for LiDAR data [Schickler and Thorpe, 2001]. One class of these approaches use the morphological erosion or an opening – represented by a morphological erosion followed by a morphological dilation. Other approaches use a watershed algorithm [Baillard et. al., 2008, Xu et. al., 2008]. In this algorithm the areas occupied by "peaks" in a digital elevation model are distinguished from each other. In the next step all such areas containing one peak are replaced by the lowest height inside and smoothed/triangulated afterward to a filled DTM.

The first class of morphological algorithms suffers from the introduction of a pre defined filter size. For urban areas such a filter radius must be larger than half of the largest minimum diameter of a building. For normal housing areas this size seldom exceeds 10 m since all rooms need day light. But for industrial buildings this filter radius can rise to 50 or more meters. On the other hand a radius of 50 m eliminates also hills up to a minimum diameter of about 100 m as in our example showing the Acropolis in Athens.

The second class fails in forests or on smooth hills. In the forests every small dip between two trees is taken as ground value whereas smooth hills get completely eliminated.

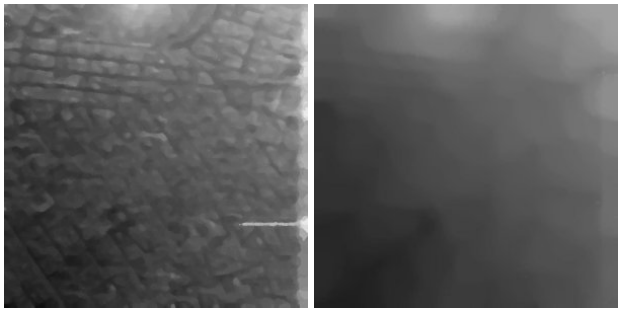

Figure 6. DSM of the Athens scene and derived DTM

For our implementation we selected the first class of algorithms using a radius of the structuring element of about 20 m. But moreover the morphological operators only work well for good situated height values. But digital surface models generated by correlation of stereo imagery suffer from many blunder values which tend to dominate the DTM generated with morphological operators. To overcome this problem a median filter which delivers a 10-%-value instead of the standard 50-%-value for substitution of the erosion and a median filter with 90-%-value for substitution of the dilation is used.

Also the median filters are calculated on a pyramid level 8 accelerating the calculation by a factor of 230 (1140 s vs. 5 s) and introducing additionally a small averaging of size 8 on the resulting DTM.

2.4 nDEM generation and object detection

The normalized DEM is generated by subtraction of the DTM from the DSM. But using this method buildings situated on steep edges of hills derive a slant on the roof from the slant of the DTM across the building. But for delineation of the buildings this effect can be neglected. After the nDEM extraction application of a Wallis filter enhances locally low buildings. Introducing a threshold of e.g. 8 m can be used for detecting high elevated objects – buildings and trees.

Applying a watershed transformation afterwards dissects the nDEM to several individual objects which can be used in turn for the subsequent modelling step.

3 EXPERIMENTAL RESULTS

3.1 Data

The very high resolution satellite imagery used for the evaluation of the described methods are three sections of two Ikonos in-orbit stereo pairs from Athens and Munich:

- 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 7, Figure 8)
- 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 9)

Figure 7. Section 1500 m × 1500 m from the Athens Stere scene, left and right stereo image showing the Acropolis (left top)

The Athens scene was already delivered as a level 1B epipolar stereo image pair. North is in the images to the right, south to the left since the orbit of the Ikonos satellite crosses Athens from north to south.

For the investigations two areas are selected from the Athens scene: The Acropolis area (Figure 7) due to the single steep hill as test for the applicability of DTM extraction algorithms and second a hilly urban area covered with many different types of buildings and a natural hill (Figure 8).

Figure 8. Section 1000 m × 1000 m from the Athens scene, also left and right stereo image (hilly urban area)

Figure 9 shows the selected sections from the original images of the Munich scene covering the city centre. This Ikonos stereo pair was acquired in forward (left image) and reverse (right image) mode due to the ordered small stereo angles. Therefore the first scan line of the left image (top line) is the northernmost line since the satellite flies from north to south. In the reverse imaging mode the first scan line is southernmost and scanning goes "reverse" of the flying path from south to north. So the topmost line in the right stereo image is the southernmost line.

Figure 9. Section 2000 m × 2000 m from the Munich scene showing the center of the city, left and right stereo image

Beside this the Munich scene was not delivered as a level 1B epipolar product due to the non standard stereo configuration ordered. So as a first preprocessing step an epipolar reprojection following [Morgan, 2004] is performed producing Figure 10.

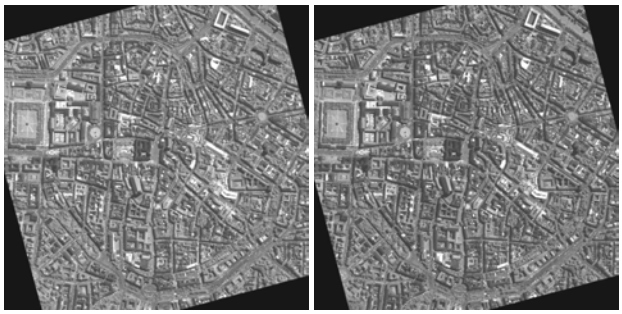

Figure 10. Section of Munich scene reprojected to epipolar geometry

3.2 Preprocessing of the raw imagery

To prepare the delivered stereo images some preprocessing is necessary. First of all the two images of the stereo pair have to be co-registered to an absolute reference or – if not available – relative to each other. This is done by transforming all images (stereo mate of the pan image and all multispectral images) to one reference (without loss of generality this is in our case the "left pan" stereo image).The co-registration of the images delivers an affine correction to the absolute reference or relative between the images. In our case the correction is only done relative and shows up that the delivered rational polynomial coefficients (RPCs, [Jacobsen et al., 2005, Grodecki et al., 2004]) with both stereo pairs fit already perfectly with less than 10 cm difference relative to each other. Absolute correction (for quality check in the case of the Munich scene) shows nevertheless a simple shift of the scene of about 4.9 m to the east and 8.4 m to the south with respect to the available reference Laser DEM.

In case of the Munich scene also a transformation to epipolar geometry as stated above is performed. The Athens scene was delivered already in epipolar geometry. An adopted epipolar transformation doesn't change the images in any case.

3.3 DSM generation

For all three sections the digital surface models are generated using the method described above with the same parameters (start window size 20 px, starting pyramid level 16, 2 layers, cost penalty for one pixel disparity 200 and cost penalty for larger disparities 1000).

The calculation is first performed using the full disparity range as shown in Figure 11, left. Afterwards using a pyramid approach which results in a speed-up of factor 3.2 (94 s vs. 29 s). In the pyramid approach the calculation of the DSM is first performed in the lowest pyramid level (in this case 16) using an area based matching with cost calculation over windows of a size of 20 px × 20 px. Going up the pyramid by a factor of two the resulting DSM is scaled up by 2 and taken as starting disparities for the calculation. In this level the calculation is done in the same way with the area based cost function but only with a window size 5 and – due to the usage of the previous DSM as offset to the actual disparities – only with a small disparity range of two times the pyramid step (=4 pixels). Only in the last step – in full resolution – the DSM calculation is performed with the original parameters (Birchfield-Tomasi cost function, no window, 2 layers …) and a

disparity range of 10 to match also thin high objects like cranes or pylons which failed to be detected in the previous pyramid levels.

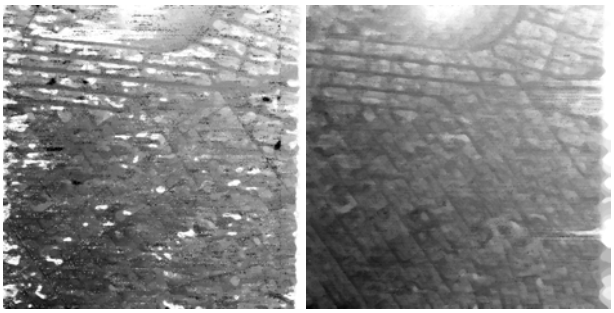

Figure 11. Resulting DSM Athens, 500 m × 500 m; left: full disparity range; right: pyramid approach with pyramid level 8

As can be seen in Figure 11 the pyramid approach (right) using only small disparity ranges in each of the pyramid levels already eliminates a large number of blunders which still exist in the full disparity range image (left).
The DSMs are calculated for all test areas with a huge range of different parameters leading to the optimal parameter set for all images (Ikonos, 11 significant bits used) as mentioned above.
A good approach for best quality is found by doing a left-right (and back) correlation followed by a right-left (and back) correlation. Each of these correlations deliver a DSM already containing detected occlusions marked as voids. Interestingly the detected occlusions differ depending on the matching direction (cf. Figure 3). So an afterward fusion of the two resulting DSMs eliminates most of the mismatches and occlusions. But this is on the cost for generating many void areas in the resulting DSM. These are filled afterwards which results in a slightly smooth DSM.

Figure 12. Computed DSM after blunder detection and filling (left) and derived DTM (right), Acropolis scene (1500 m × 1500 m)

The DSM derived from the Acropolis section of the Athens scene after adaptive median filling and blunder detection is shown together with the derived DTM in Figure 12. Additionally a mask of high objects (filtered nDEM with threshold of 5 meters) and the vegetation mask (NDVI with a threshold of 140) can be derived.

3.4 DTM extraction

After ortho projection of the generated DSM the described DTM extraction using the 10-%-median filter followed by a 15-%-median on an image reduced by a factor of 8 and scaling up afterwards is performed. The results for the Athens and the Acropolis section are shown in Figure 6 and Figure 12 respectively. In the Munich scene the DTM is nearly flat.

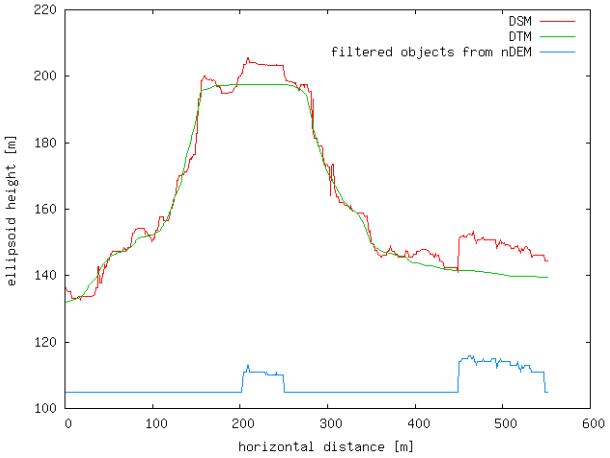

Figure 13. Profile north to south across the Acropolis showing the calculated DSM (red), the derived DTM (green) and the filtered and extracted buildings (blue at bottom)

4 RESULTS

4.1 Quality

For quality control a LiDAR reference DEM (Laser DEM) was only available for the Munich scene. The results are shown in Figure 14 and Figure 15.

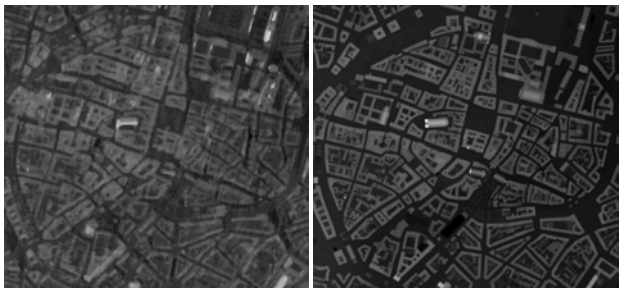

Figure 14. Munich Scene, Section 1400 m × 1300 m, left calculated DSM, right reference Laser-DEM (with no trees), both UTM Zone 32, WGS84 ellipsoid heights

Creating statistics on the differences between the calculated DSM and the Laser DEM gives Figure 15 with a height binning of 1 m.

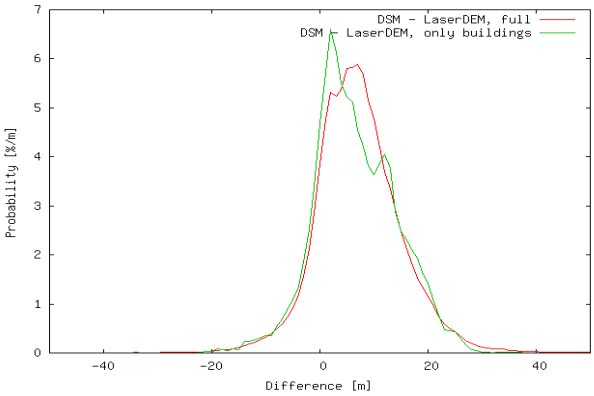

Figure 15. Statistics of differences between calculated DSM and Laser DEM; green: building areas; red: full differences (including trees)

It can be seen that the calculated DSM is mainly a little bit above the Laser DSM due to the smoothing effects at walls, filled up narrow streets and court yards. Moreover the Laser DEM does not include any trees and so the statistics calculated over the full scene shows a significant increase of height differences.
The average offset of the generated DSM relative to the Laser DEM is 6.6 m with a standard deviation of about 7.6 m. The overall DEM – including the trees – shows a shift of 7.3 m and a standard deviation of 8.1 m relative to the Laser DEM.

4.2 Processing times

The processing time using the pyramid approach with only small disparity ranges results in a speed-up of about 7 times (254.5 s vs. 36.6 s). Also in contrast to the classical ATE approach with selecting interest points, area based matching (see [Lehner and Gill, 1992]) and interpolating the DSM the speed-up is about a factor of 6.5.

Starting pyramid level	run time
no pyramid	254.5 s
2	350.8 s
4	61.7 s
8	48.0 s
16	36.6 s
32	34.0 s

(all calculated with 512 px × 512 px images)

Calculating DSMs with pyramid level 16 but different edge lengths from 512 to 2000 using the Munich scene shows a time dependency of O(W H) for width W and height H of the image.

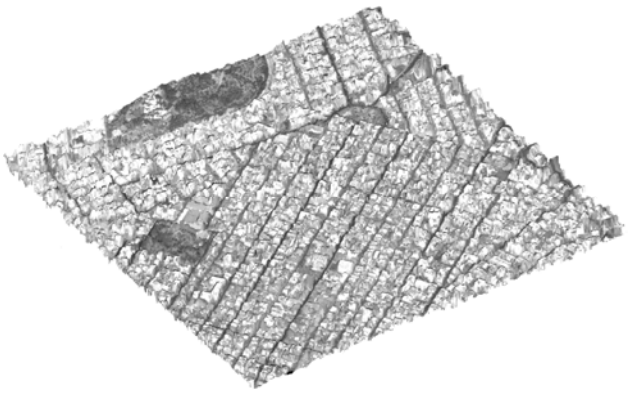

Figure 16. 3D projection of the Athens DEM overlaid with the ortho image

5 CONCLUSION AND OUTLOOK

In this paper an improved fusion approach of the DSM generation from VHR satellite imagery is presented. In contrast to classical approaches the DSM generation is performed by a dense stereo approach processed in an image pyramid. Intrinsic to the process is the automatic detection of occlusions and blunders. A subsequential fusion and interpolation process of the derived DSMs eliminates additionally small blunders from the resulting filled DSM. After orthorectification a DTM is derived which in turn allows the calculation of the normalized DEM and the extraction of high elevated objects. Also via calculation of the NDVI image from the multi spectral stereo images a vegetation mask can be extracted.

Taking these two masks in account the trees (high vegetation) and buildings (high, non vegetation objects) can be extracted and their outlines straightened by calculating main orientations in different pyramid levels from the produced DSM and the original stereo imagery.

REFERENCES

Birchfield, S. and C. Tomasi, C., 1998. A Pixel Dissimilarity Measure That is Insensitive To Image Sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(4):401-406, April 1998.

Baillard, C., 2008. A hybrid method for deriving DTMs from urban DEMs. ISPRS J., 29 (3), 21. ISPRS Congress, July 2008, Beijing

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).

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. In: IEEE transactions on pattern analyses and machine intelligence, 30 (2), Feb. 2008.

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.

Lehner, M. and d'Angelo, P. and Müller, R. and Reinartz, P, 2008. Stereo Evaluation of CARTOSAT-1 Data Summary of DLR Results During CARTOSAT-1 Scientific Assessment Program., ISPRS J., 29 (3), p.1295 ff, 21. ISPRS Congress, July 2008, Beijing

Morgan, M. F., 2004. Epipolar resampling of linear array scanner scenes. Phd Thesis, Geomatics Engineering, University of Calgary, http://hdl.handle.net/1880/41819

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

Pentenrieder, C., 2008. Analyse und Vergleich von 3D-Stereo-Verfahren für hochauflösende Satellitenbilder. Diploma theseis, Hochschule für angewandte Wissenschaften, München

Scharstein, D. and Szeliski, R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision, 47(1/2/3):7-42, April-June 2002. Microsoft Research Technical Report MSR-TR-2001-81, November 2001. Middlebury stereo vision page. http://vision.middlebury.edu/stereo. (accessed 04/2008).

Schickler, W. and Thorpe, A., 2001. Surface estimation based on LIDAR. ASPRS Conference, St. Louis, Missouri, USA, April 2001

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

Xu, F. and Woodhouse, N. and Xu, Z. and Marr, D. and Yang, X. and Wang, Y., 2008. Blunder elimination techniques in adaptive automatic terrain extraction. ISPRS J., 29 (3), 21. p.1139 ff, ISPRS Congress, July 2008, Beijing