

DSM ACCURACY EVALUATION FOR THE ISPRS COMMISSION I IMAGE MATCHING BENCHMARK

G. Kuschka^a, P. d'Angelo^a, R. Qin^b, D. Poli^c, P. Reinartz^a, D. Cremers^d

^a German Aerospace Center (DLR), Remote Sensing Technology Institute, 82230 Wessling, Germany -
(georg.kuschka,pablo.angelo,peter.reinartz)@dlr.de

^b Singapore ETH Center, Future Cities Laboratory, CREATE Tower Singapore 138602 - rqin@student.ethz.ch

^c Terra Messflug GmbH, Eichenweg 42, 6460 Imst - danipolix@gmail.com

^d TU Munich, Germany - cremers@tum.de

Commission VI, WG VI/4

KEY WORDS: 3D Stereo Reconstruction, DEM/DSM, Matching, Accuracy, Benchmark

ABSTRACT:

To improve the quality of algorithms for automatic generation of Digital Surface Models (DSM) from optical stereo data in the remote sensing community, the Working Group 4 of Commission I: Geometric and Radiometric Modeling of Optical Airborne and Spaceborne Sensors provides on its website <http://www2.isprs.org/commissions/comm1/wg4/benchmark-test.html> a benchmark dataset for measuring and comparing the accuracy of dense stereo algorithms. The data provided consists of several optical spaceborne stereo images together with ground truth data produced by aerial laser scanning. In this paper we present our latest work on this benchmark, based upon previous work.

As a first point, we noticed that providing the abovementioned test data as geo-referenced satellite images together with their corresponding RPC camera model seems too high a burden for being used widely by other researchers, as a considerable effort still has to be made to integrate the test data's camera model into the researchers' local stereo reconstruction framework. To bypass this problem, we now also provide additional rectified input images, which enable stereo algorithms to work out of the box without the need for implementing special camera models. Care was taken to minimize the errors resulting from the rectification transformation and the involved image resampling.

We further improved the robustness of the evaluation method against errors in the orientation of the satellite images (with respect to the LiDAR ground truth). To this end we implemented a point cloud alignment of the DSM and the LiDAR reference points using an Iterative Closest Point (ICP) algorithm and an estimation of the best fitting transformation. This way, we concentrate on the errors from the stereo reconstruction and make sure that the result is not biased by errors in the absolute orientation of the satellite images.

The evaluation of the stereo algorithms is done by triangulating the resulting (filled) DSMs and computing for each LiDAR point the nearest Euclidean distance to the DSM surface. We implemented an adaptive triangulation method minimizing the second order derivative of the surface in a local neighborhood, which captures the real surface more accurately than a fixed triangulation. As a further advantage, using our point-to-surface evaluation, we are also able to evaluate non-uniformly sampled DSMs or triangulated 3D models in general. The latter is for example needed when evaluating building extraction and data reduction algorithms.

As a practical example we compare results from three different matching methods applied to the data available within the benchmark data sets. These results are analyzed using the above mentioned methodology and show advantages and disadvantages of the different methods, also depending on the land cover classes.

1. INTRODUCTION

Given the rising number of available optical satellites and their steadily improving ground sampling resolution, as well as advanced software algorithms, large-scale 3D stereo reconstruction from optical sensors is increasingly getting a widespread method to cost-efficiently create accurate and detailed digital surface models (DSMs). Their manifold usage includes for example monitoring of inaccessible areas, change detection, flood simulation / disaster management, just to name a few. Due to the amount of data, and with the prospect of a real-time global coverage by a network of optical satellites in the near future, the algorithms used for 3D stereo reconstruction have work fully automatic.

A natural engineering question for the development of these 3D stereo reconstruction algorithms is the availability of a common objective test-bed environment, comparing the stereo results with a known ground truth. In many computer vision related areas, such benchmarks boosted the development of their respective fields, as they directly allow to compare their numerical accuracy. In the area of 3D stereo reconstruction, the most prominent benchmarks

are the Middlebury stereo benchmark (Scharstein and Szeliski, 2002) and the KITTI stereo benchmark (Geiger et al., 2012). Despite the challenging data sets, especially so the KITTI benchmark, their image acquisition is focused on ground based indoor and outdoor scenarios (e.g. automotive). So even if some of the most common stereo problems are to some degree existent in the data sets (sensor oversaturation due to glaring sunshine, textureless areas and repetitive textures, low-contrast shadow areas, occlusions), the nature of the image acquisition and their sensor properties are quite different to the ones used in optical satellites. For aerial imagery an additional benchmark, EuroSDR (Haala, 2013), is available, which unfortunately does neither provide a comparison of the submitted algorithms and their results, neither is the ground truth data (LiDAR points) publicly available.

To address the aforementioned drawbacks, a stereo benchmark for dense stereo matching of optical satellite images was introduced by (Reinartz et al., 2010). In this work we build upon this benchmark and improve the evaluation procedure, provide more detailed results in different types of urban and rural sub-areas, upload the submissions with the applicants' consent to the corre-

sponding web page and thereby enable a objective online comparison. We further simplified the first time usage of the data by providing rectified versions of the image data sets, thus enabling participants to also run tests without having to care about the camera model. And additionally, we provide own results for two, now standard, dense stereo matching methods. These won't be put into the official accuracy ranking, as we deem it unethical to provide a benchmark dataset, take care of the evaluation of its submissions, and at the same time participate in it.

2. BENCHMARK DATA

The publicly available optical satellite datasets used for the benchmark were introduced in (Reinartz et al., 2010) and cover three characteristic sub-regions near Barcelona, Spain (see Figure 1). To further investigate the accuracy of optical stereo reconstruction methods in complex urban areas, one of these sub-regions (Terrassa) was manually classified into typical city areas like industrial, residential, ... (see Figure 2). For more detailed information on the satellite image acquisition properties as well as the reference data see (d'Angelo and Reinartz, 2011) and (Reinartz et al., 2010). All datasets and reference data is and will be provided at (ISPRS Satellite Stereo Benchmark, 2014), were all evaluations are and will be listed as well.

2.1 Datasets

Three datasets are part of the benchmark:

- Subsets of a Cartosat-1 Stereo pair, acquired in February 2008. It consists of a nadir viewing (-5°) and a forward viewing ($+27^\circ$) scene. The ground sample distance is 2.5m.
- Worldview-1, one L1B stereo scene acquired in August 2008. One Nadir image (-1.3°) and one off-nadir scene (33.9°), with ground sample distances of 0.5 and 0.76 meters.
- A Pleiades triplet, acquired on 8th of January 2013, with the following off nadir angles: forward image: 18° , nadir image: 8° , and backward image: 17° .

A RPC based block adjustment was performed, independently for each test area and sensor. GCPs derived from the ortho images and LiDAR point cloud are used as reference, ensuring a good coregistration of the DSMs with the reference LiDAR point cloud. Both ortho images and LiDAR point clouds are provided by the Institut Cartogràfic de Catalunya (ICC) - see Section 2.2. For Cartosat-1, a 6-parameter affine RPC correction was used. For Worldview-1 and Pleiades, a row and column shift was sufficient. The corrections were applied to the RPCs, resulting in a good relative and absolute orientation of all datasets.

Benchmark participants can directly use the RPCs provided with the data without any further adjustment or correction with GCPs.

2.2 Reference Data

The reference data consist of color orthoimages with a spatial resolution of 50 cm as well as an airborne laser scanning point cloud (first pulse and last pulse) with approximately 0.3 points per square meter. The LiDAR points were acquired end of November 2007.

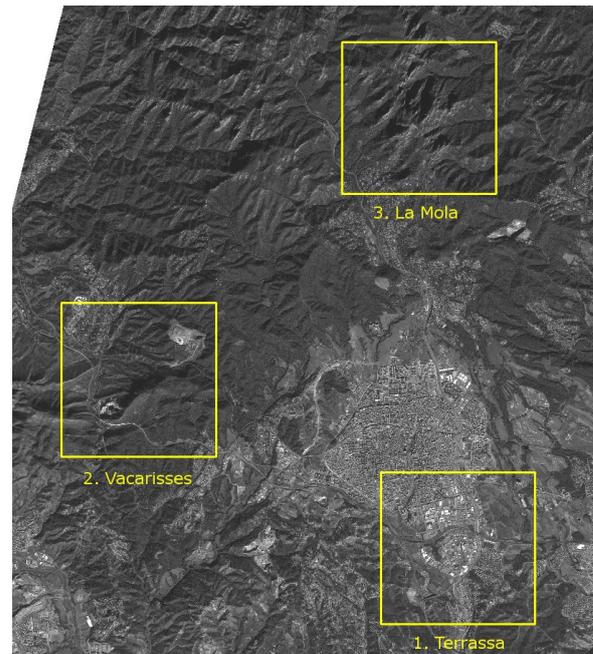


Figure 1: Cartosat-1 image showing the three 4 km×4 km test areas: Terrassa, Vacarisses, La Mola



Figure 2: Cartosat-1 image showing the subareas in the urban 'Terrassa' area. Blue: Industrial; Orange: Residential; Yellow: City; Purple: Bridges; Green: Open fields; Red: Changes

2.3 Epipolar rectification

For enabling users to quickly test their stereo methods on our benchmark, we additionally provide rectified versions of the satellite images, allowing a 3D reconstruction by computing their disparity maps without having to deal with the RPC camera model. It should be noted that, although we provide a Windows binary for converting the computed disparity maps back into a UTM based DSM, the original satellite images and their RPC's should be used to create DSMs with highest quality. The two ways of generating a DSM, either via RPCs or using the rectified images, are shown in Figure 3.

The rigorous model of the linear push-broom sensors provides

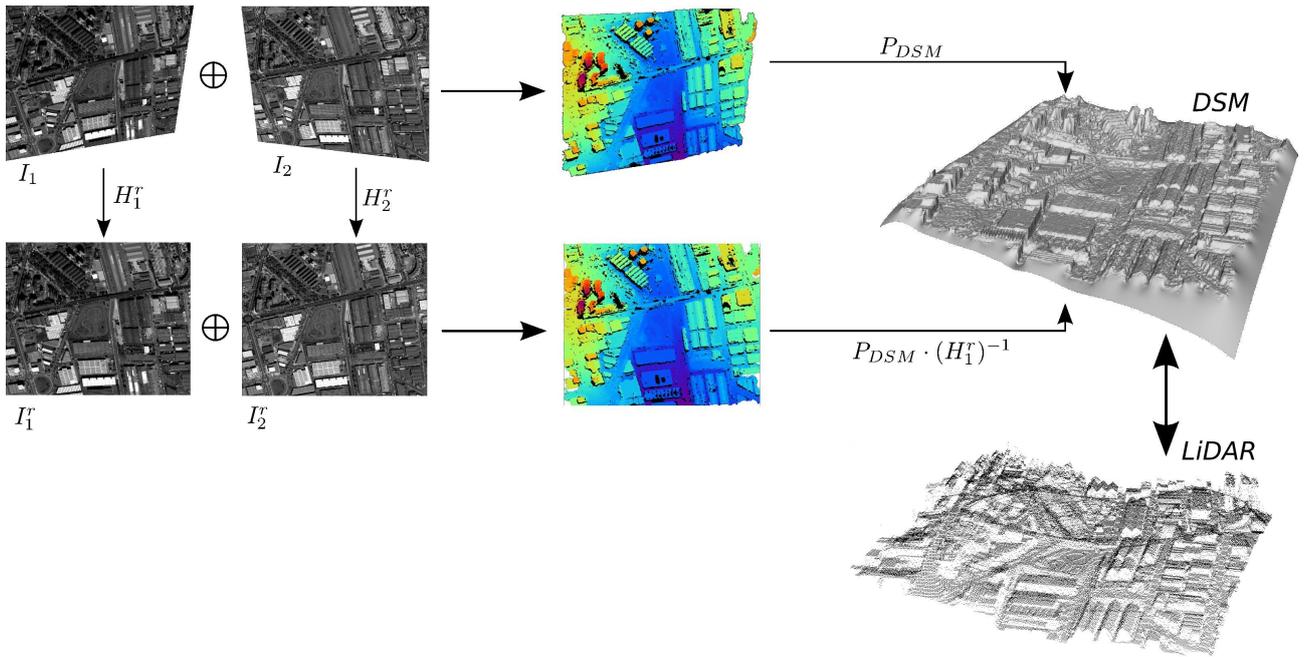


Figure 3: Two options to create the DSM, either with original satellite images and RPC camera model or using the rectified images. The triangulated DSM is then evaluated against the LiDAR ground truth

quasi-epipolar geometry for well-orientated images, where the search space of a point can be constrained to a curve in the search image. Due to the high stability of the spaceborne platform, this could be approximated as a straight line for data covering a relatively small area (Morgan et al., 2004), like in the case of our benchmark dataset. Therefore, the stereo images can be digitally rectified in a way that the matching correspondences are located on the same row of the rectified image grids, thus simplifying the matching problem and not needing the camera geometry anymore to produce a disparity map (Hartley and Zisserman, 2003).

We adopt a two-step strategy for computing the rectified images, similar to the one proposed by (Wang et al., 2011). The basic idea of this approach is to first project the original images to a common reference plane parallel to the ground, and then applying a 2D affine transformation to align the equivalent epipolar lines (EEL) to the same sampling row in the image grid. In most of the cases the RPC (Rational Polynomial Coefficient) are provided as a backward projection, namely from the object space to the image space, defined as

$$x = \frac{Samp_{Num}(X, Y, Z)}{Samp_{Den}(X, Y, Z)}, y = \frac{Line_{Num}(X, Y, Z)}{Line_{Den}(X, Y, Z)} \quad (1)$$

while the forward projection is computed iteratively by giving the height value. To rectify the projected images, we first need to find the pairs of EEL, which carry all the correspondence for each other. Simply put, the correspondences of points on the epipolar line in the left image can be found in its EEL on the right image. A pair of EEL can be computed with the following procedure: Given a point in the left image, we first compute its corresponding epipolar line in the right image (Zhao et al., 2008), and then based on the central point of this epipolar line, find its corresponding epipolar line back to the left image. These two lines are equivalent and contain the matching correspondences for each other. Given pairs of EEL, the transformation parameters for aligning these pairs of the EEL can be estimated by minimizing their errors in the vertical direction. We employ a 2D rigid transformation for the image rectification, which consists of rotation and translation. In practice, the 2D rigid transformation can be calcu-

lated by estimating the average rotation angles and y-axis offsets of the EEL pairs in a step-wise manner. The rotation angle θ_L for the left image is computed as:

$$\theta_L = \frac{1}{n} \sum_i^n \arccos(k_{l,i}) \quad (2)$$

where $k_{l,i}$ is the slope of the epipolar line i in the left image, and n is the number of considered EELs. The angle θ_R for the right image can be computed similarly. After applying the rotation to both images, the vertical shift t_l and t_r can be estimated by taking the mean value of the vertical shifts of the rotated EEL:

$$t_l = \frac{1}{n} \sum_i^n \frac{(y_{r,i} - y_{l,i})}{2}$$

$$t_r = -t_l \quad (3)$$

where $y_{l,i}$ and $y_{r,i}$ are the vertical coordinates of the rotated EELs on the left and right images, respectively. The transformation from the original images I to the rectified image I_{rect} can be then described by the coordinate transformation

$$p_{rect} = A \begin{bmatrix} p \\ 1 \end{bmatrix} = [R_\theta \ T] \begin{bmatrix} P \\ 1 \end{bmatrix} \quad (4)$$

where A is the affine transformation matrix containing the rotation matrix $R_\theta \in \mathbb{R}^{2 \times 2}$ with respect to the angle θ , plus the vertical translation vector $T = [0 \ t]^T$.

(Wang et al., 2011) have proven that these pairs of EEL have small residuals in terms of their agreement of the slope, and the final y parallax is within 0.5 pixels for most of the sensors. We derived the 2D rigid transformation parameters based on three pairs of EELs, which were chosen evenly across the vertical direction of the image.

3. EVALUATION PROCEDURE

The automatic evaluation procedure requires the submitted DSMs to be projected into UTM Zone 31 North and sampled with two times the GSD (5m for Cartosat-1 and 1m for Worldview-1). Resulting holes need to be filled, otherwise a B-Spline based interpolation method is used to fill remaining holes (Lee et al., 1997). In a first step, to ensure that the evaluation is not biased by errors in the orientation of the satellite images, the (x, y, z) -shift of the DSMs with respect to the reference point clouds is computed, using an Iterative Closest-Point Algorithm (Besl and McKay, 1992) and the 3D-3D least-squares registration method of (Umeyama, 1991).

The final evaluation step is done by computing for each LiDAR point of the reference point cloud its distance to the meshed DSM surface. The LiDAR point is being projected onto the 2D grid of the corresponding DSM, falling between four incident pixels (x, y) , $(x + 1, y)$, $(x + 1, y + 1)$, $(x, y + 1)$. Instead of comparing the LiDAR height value with the interpolated height value of the DSM, we use a more robust distance measure and compute the minimum distance of the LiDAR point to the triangles created by the two possible triangulations of the DSM surface as shown in Figure 4.

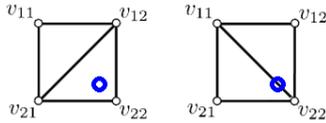


Figure 4: The two possible triangulations of a square involving four incident pixels v_{ij} and the projected LiDAR point in blue.

As result, each LiDAR point from the reference data is associated with its distance to the computed DSM and the overall mean absolute error (MAE) per test area is computed. For the different sub-areas of the dataset *Terrassa* (as shown in Figure 2), masks are applied to only evaluate the error inside these regions. For the final ranking of the algorithm's accuracy, the average MAE over all sub-areas is computed (compare Figure 6).

To further give an intuitive feedback of the accuracy, locating areas where algorithms perform quite well or bad, we also provide visual error maps as show in Figure 5. For this purpose we project the LiDAR points (and their distances to the computed DSM respectively) onto a 2D grid, having the same geo-reference and ground sampling distance as the computed DSMs.

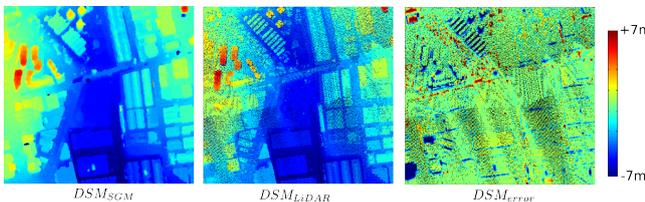


Figure 5: Evaluation of a sample area using the described approach

4. DSM GENERATION

To provide first example results, we processed the benchmark data with two standard dense stereo matching methods, described in this section.

Let the image space of the reference image I_1 be denoted as $\Omega \subset \mathbb{R}^2$. For every pixel $\mathbf{x} = (x, y)^T \in \Omega$ and every height hypothesis $\gamma \in \Gamma = [\gamma_{min}, \gamma_{max}]$, we compute a matching cost

$C(\mathbf{x}, \gamma)$ with respect to a second image I_2 . The matching cost function is defined as

$$C(\mathbf{x}, \gamma) = Diff(Desc(I_1, \mathbf{x}), Desc(I_2, \pi(\mathbf{x}, \gamma))) \quad (5)$$

with $Desc$ being a local image descriptor, π the function projecting a pixel $\mathbf{x} \in I_1$ to image I_2 by using the height γ , and $Diff$ a function to measure the difference of the aforementioned image descriptors.

As local image descriptors $Desc$ we used the Census transform (Zabih and Woodfill, 1994) with a 9×7 window, measuring their difference by the Hamming distance of their respective bit strings.

In the resulting three-dimensional $(\Omega \times \Gamma)$ disparity space image (DSI), also called cost cube, we now search for a functional $u(\mathbf{x})$ (the disparity map), which minimizes the energy function arising from the matching costs (called data term E_{data}), plus additional regularization terms. As the data term is prone to errors due to some incorrect and noisy measurements, one often needs some smoothness constraints E_{smooth} , forcing the surface of the disparity map to be locally smooth.

$$u(\mathbf{x}) = \underset{u}{\operatorname{argmin}} \left\{ \int_{\Omega} E_{data} + E_{smooth} \, d\mathbf{x} \right\} \\ = \underset{u}{\operatorname{argmin}} \left\{ \int_{\Omega} \lambda \cdot C(\mathbf{x}, u(\mathbf{x})) + \nabla u(\mathbf{x}) \, d\mathbf{x} \right\} \quad (6)$$

This energy is non-trivial to solve, since the smoothness constraints are based on gradients of the disparity map and therefore cannot be optimized pixelwise anymore. Various approximations for this NP-hard problem are existing, e.g. Semi-global Matching (Hirschmüller, 2005), energy minimization via graph cuts (Boykov et al., 2001) or minimization of Total Variation (Pock et al., 2008) - just to name three examples. When the matching costs C and the smoothness penalties ∇u are normed, one can steer the smoothness of the resulting height map using a scalar balancing parameter λ without the need to choose it manually, depending on the height values of the given data and there possible gradients respectively.

4.1 Semi-global Matching (SGM)

In the work of (Hirschmüller, 2005), the energy functional of Equation 6 is approximated as follows

$$E(u) = \sum_{\mathbf{x}} C(\mathbf{x}, u(\mathbf{x})) + \\ \sum_{\mathbf{p} \in N_{\mathbf{x}}} P_1 \cdot T[|u(\mathbf{x}) - u(\mathbf{p})| = 1] + \\ \sum_{\mathbf{p} \in N_{\mathbf{x}}} P_2 \cdot T[|u(\mathbf{x}) - u(\mathbf{p})| > 1] \quad (7)$$

The first term is the matching cost, the second and third term are penalties for small and large disparity discontinuities between neighboring pixels $\mathbf{p} \in N_{\mathbf{x}}$. The parameters P_1 and P_2 serve as balancing parameter between the data term and the smoothness terms (similar to λ in Equation 6). The key idea is now not to solve the intractable global 2D problem, but to approximate it by combining 1D solutions from different directions \mathbf{r} , solved via dynamic programming

$$u(\mathbf{x}) = \underset{u}{\operatorname{argmin}} \left\{ \sum_{\mathbf{r}} L_{\mathbf{r}}(\mathbf{x}, u(\mathbf{x})) \right\} \quad (8)$$

For details about how the aggregated costs $L_{\mathbf{r}}$ are computed, the reader is referred to (Hirschmüller, 2005). In our results (Section

5.), the penalties were set to $P_1 = 0.4$ and $P_2 = 0.8$ (with the raw costs normed to $[0, 1]$).

4.2 Minimizing Total Variation (TV)

In a second approach we minimize the Total Variation of Equation 6 using the primal-dual algorithm of (Pock et al., 2008).

$$u(\mathbf{x}) = \underset{u}{\operatorname{argmin}} \left\{ \int_{\Omega} \lambda \cdot C(\mathbf{x}, u(\mathbf{x})) + \nabla u(\mathbf{x}) \, d\mathbf{x} \right\}$$

$$\approx \underset{u,a}{\operatorname{argmin}} \left\{ \int_{\Omega} \lambda \cdot C(\mathbf{x}, a(\mathbf{x})) + \nabla u(\mathbf{x}) + \frac{1}{2\theta}(u - a)^2 \, d\mathbf{x} \right\} \quad (9)$$

In (Steinbruecker et al., 2009), a quadratic relaxation between the convex regularizer $\nabla u(\mathbf{x})$ and the non-convex data term $C(\mathbf{x}, u(\mathbf{x}))$ was proposed for minimizing a Total Variation based optical flow energy functional and (Newcombe et al., 2011) used a similar approach for image driven and TV-based stereo estimation. We build upon these ideas and split the stereo problem from Equation 6 into two subproblems and, using quadratic relaxation, couple the convex regularizer $R(\mathbf{u})$ and non-convex data term $C(\mathbf{u})$ through an auxiliary variable \mathbf{a} :

$$E = \int_{\Omega} R(\mathbf{u}) + \lambda \cdot C(\mathbf{a}) + \frac{1}{2\theta}(\mathbf{u} - \mathbf{a})^2 \, d\mathbf{x} \quad (10)$$

By iteratively decreasing $\theta \rightarrow 0$, the two variables \mathbf{u} , \mathbf{a} are drawn together, enforcing the equality constraint $\mathbf{u} = \mathbf{a}$. While the regularization term is convex in u and can be solved efficiently using a primal-dual approach for a fixed auxiliary variable a , the non-convex data term can be solved point-wise by an exhaustive search over a set of discretely sampled disparity values. This process is done alternatingly in an iterative way. For further details about the minimization procedure, the reader is referred to the work of (Pock et al., 2008), (Newcombe et al., 2011), (Kuschk and Cremers, 2013).

In our results (Section 5.), the impact of the data term was set to $\lambda = 0.1$ (with the raw costs normed to $[0, 1]$).

5. RESULTS

We provide initial results for two dense stereo matching methods, which won't be put into the official accuracy ranking. These two algorithms were described in the former section: SGM - Section 4.1 and a Variational approach - Section 4.2. Except for the stereo algorithms themselves, all other parameters were fixed to the same settings (e.g. post-processing steps like median filtering and interpolation of holes due to UTM projection or the Census transform as image matching cost function).

No exhaustive parameter tuning was performed, and the parameters for the impact of the smoothness term were chosen in a way to provide a similar smoothness of the resulting DSMs. As image matching cost function, a simple Census Transform was chosen, see Section 4..

6. CONCLUSIONS

In this work we improved and extended the existing ISPRS Stereo Benchmark. The submitted DSMs are evaluated for all three test areas and with respect to different land cover types as well. All

numerical and visual results will be displayed with the applicants consent on the corresponding web page and thereby enable a objective online comparison for the different algorithms. All datasets and reference data are publicly available and we heartily invite researchers in the field of 3D reconstruction from remote sensing imagery to evaluate their algorithms with the proposed benchmark. In a next step we plan to do a thorough evaluation of the state of the art image matching cost functions and the optimization methods for the energy functional to minimize.

ACKNOWLEDGMENT

The authors would like to thank Euromap for providing the Cartosat-1 data, Digital Globe for the Worldview-1 data and the Institut Cartogràfic de Catalunya for the reference data.

Rank	Algorithm	Worldview										Cartosat							Details	
		Average	La Mola	Vacarisses	Terrassa	Bridge	City	Field	Industrial	Residential	Average	La Mola	Vacarisses	Terrassa	Bridge	City	Field	Industrial		Residential
X	TV	1.87	2.35	2.22	1.04	1.17	1.95	0.48	1.01	1.32	2.81	3.67	2.71	2.04	2.32	3.13	1.42	2.00	2.43	Details
X	SGM	2.22	2.64	2.90	1.11	1.18	1.91	0.60	1.12	1.36	3.36	4.52	3.35	2.22	2.67	3.41	1.50	2.23	2.68	Details

Notes:

- Submitted DSMs MUST not contain background pixel-values in the evaluated areas
- Error metrics are Mean Absolute Error (MAE) in meter
- Average = Average MAE of "La Mola", "Vacarisses", "Terrassa"
- Errors maps are showing the absolute error between [-10..10m], scaled to grayvalues [0..250]
 Errors outside this interval are truncated

Figure 6: Evaluation results (22.10.2014) for two stereo algorithms: SGM and TV - see Section 4. for details.

REFERENCES

Besl, P. and McKay, N., 1992. A Method for Registration of 3-D Shapes. *IEEE Transactions on pattern analysis and machine intelligence* 14(2), pp. 239–256.

Boykov, Y., Veksler, O. and Zabih, R., 2001. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Transactions on pattern analysis and machine intelligence* pp. 1222–1239. Graph Cuts.

d'Angelo, P. and Reinartz, P., 2011. Semiglobal Matching Results on the ISPRS Stereo Matching Benchmark. Vol. 3819, pp. 79–84.

Geiger, A., Lenz, P. and Urtasun, R., 2012. Are we Ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE*, pp. 3354–3361.

Haala, N., 2013. The Landscape of Dense Image Matching Algorithms. In: *Photogrammetric Week 2013*, Wichmann Verlag, pp. 271–284.

Hartley, R. and Zisserman, A., 2003. *Multiple View Geometry in Computer Vision*. Cambridge Univ Pr.

Hirschmuller, H., 2005. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 2, IEEE*, pp. 807–814.

ISPRS Satellite Stereo Benchmark, 2014. <http://www2.isprs.org/commissions/commi/wg4/benchmark-test.html> [Online; Status 15-Oct-2014].

Kuschk, G. and Cremers, D., 2013. Fast and Accurate Large-scale Stereo Reconstruction using Variational Methods. In: *ICCV Workshop on Big Data in 3D Computer Vision*, Sydney, Australia.

Lee, S., Wolberg, G. and Shin, S., 1997. Scattered Data Interpolation with Multilevel B-Splines. *Visualization and Computer Graphics, IEEE Transactions on* 3(3), pp. 228–244.

Morgan, M., Kim, K., Jeong, S. and Habib, A., 2004. Epipolar geometry of linear array scanners moving with constant velocity and constant attitude. *ISPRS, Istanbul, Turkey*.

Newcombe, R., Lovegrove, S. and Davison, A., 2011. DTAM: Dense Tracking and Mapping in Real-Time. In: *Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE*, pp. 2320–2327.

Pock, T., Schoenemann, T., Graber, G., Bischof, H. and Cremers, D., 2008. A Convex Formulation of Continuous Multi-Label Problems. *Computer Vision - ECCV 2008* pp. 792–805.

Reinartz, P., d'Angelo, P., Krauß, T., Poli, D., Jacobsen, K. and Buyuksalih, G., 2010. Benchmarking and Quality Analysis of DEM Generated from High and Very High Resolution Optical Stereo Satellite Data. Vol. 3818, pp. 1–6.

Scharstein, D. and Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International journal of computer vision* 47(1), pp. 7–42.

Steinbruecker, F., Pock, T. and Cremers, D., 2009. Large Displacement Optical Flow Computation without Warping. In: *Computer Vision, 2009 IEEE 12th International Conference on, IEEE*, pp. 1609–1614.

Umeyama, S., 1991. Least-Squares Estimation of Transformation Parameters between two Point Patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 13(4), pp. 376–380.

Wang, M., Hu, F. and Li, J., 2011. Epipolar resampling of linear pushbroom satellite imagery by a new epipolarity model. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(3), pp. 347–355.

Zabih, R. and Woodfill, J., 1994. Non-parametric Local Transforms for Computing Visual Correspondence. *Computer Vision - ECCV 1994* pp. 151–158. census transform, rank transform.

Zhao, D., Yuan, X. and Liu, X., 2008. Epipolar line generation from ikonos imagery based on rational function model. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37(B4), pp. 1293–1297.