EXPLORING THE APPLICABILITY OF SEMI-GLOBAL MATCHING FOR SAR-OPTICAL STEREOGRAMMETRY OF URBAN SCENES

H. Bagheri [a], M. Schmitt [a], P. d'Angelo [b], XX. Zhu[a,b]

[a] Signal Processing in Earth Observation, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany-(hossein.bagheri, m.schmitt)@tum.de
[b] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany-(xiao.zhu, pablo.angelo)@dlr.de

Commission II, WG II/2

KEY WORDS: TerraSAR-X, WorldView-2, dense matching, semi-global matching (SGM), data fusion, SAR-optical stereogrammetry

ABSTRACT:

Nowadays, a huge archive of data from different satellite sensors is available for diverse objectives. While every new sensor provides data with ever higher resolution and more sophisticated special properties, using the data acquired by only one sensor might sometimes still not be enough. As a result, data fusion techniques can be applied with the aim of jointly exploiting data from multiple sensors. One example is to produce 3D information from optical and SAR imagery by employing stereogrammetric methods. This paper investigates the application of the semi-global matching (SGM) framework for 3D reconstruction from SAR-optical image pairs. For this objective, first a multi-sensor block adjustment is carried out to align the optical image with a corresponding SAR image using an RPC-based formulation of the imaging models. Then, a dense image matching, SGM is implemented to investigate its potential for multi-sensor 3D reconstruction. While the results achieved with Worldview-2 and TerraSAR-X images demonstrate the general feasibility of SAR-optical stereogrammetry, they also show the limited applicability of SGM for this task in its out-of-the-box formulation.

1. INTRODUCTION

Today, a growing number of satellites equipped with various kinds of sensors provide remotely-sensed images of our planet. Each kind of sensor has its own distinct properties regarding, e.g., wavelength, resolution, accuracy and coverage. As an example, the TanDEM-X mission provided high-resolution bistatic SAR images for the whole earth. Moreover, a large coverage of landmasses is enabled by the Sentinel-1 and 2 missions. The Sentinel-1 mission provides a global, cloud-free medium-resolution SAR dataset that can well be used for large-scale terrain and deformation reconstruction, while Sentinel-2 provides easy-to-interpret multi-spectral data that is well-suited for land-use/land-cover mapping tasks being heavily affected by cloud coverage. Furthermore, high and very high-resolution images are acquired by the modern generation of optical sensors like WorldView-2, 3 and 4. As a result, large archives of satellite imagery acquired by different sensors are available and will not stop to grow in the future. Data fusion can be applied for integrating datasets with different specifications to enhance information extraction by beneficially combining the individual sensors properties (Schmitt and Zhu, 2016).

Of particular interest in that regard is the fusion of optical and SAR imagery for different purposes (Schmitt et al., 2017). One of the purposes is 3D reconstruction by SAR-optical stereogrammetry, which also rectifies the optical imagery with modern SAR sensors such as TerraSAR-X. The geolocation accuracy of basic products of WorldView-2 imagery is 5m while the geolocolaization accuracy of high resolution spotlight TerraSAR-X imagery is better than the size of a resolution cell in both azimuth and range directions (Eldhuset and Weydahl, 2011, DigitalGlobe, 2018). Since optical imagery suffers from a poor absolute positioning accuracy in comparison to SAR, matching of optical images to SAR data takes can improve the absolute accuracy of optical acquisitions.

In this context, the main idea of this paper is to investigate the applicability of the semi-global matching (SGM) algorithm for SAR-optical stereogrammetry, and to design a framework for accomplishing this task. The output of this study identifies both potentials and limitations of out-of-the-box SGM applied SAR and optical images to produce 3D data over urban areas.

2. STEREOGRAMMETRIC 3D RECONSTRUCTION FROM SAR-OPTICAL IMAGE PAIRS

2.1 Semi-Global Matching for SAR-Optical Stereogrammetry

The main stage of stereogrammetric 3D reconstruction is to carry out a dense image matching algorithm to generate a disparity map. Dense matching can be implemented in a local or a global manner. Global dense matching establishes an energy functional to find the disparity map that makes the matching process more robust against noise than local matching methods. In this study, a global dense image matching procedure is implemented for stereo SAR-optical images to produce the corresponding disparity map. For this purpose, SGM with two different cost functions, namely Mutual Information (MI) and Census, as well as the weighted sum of both cost functions, is implemented for dense matching of TerraSAR-X and WorldView-2 imagery.

The main property of SGM is to find a disparity map with the minimum cost function over the whole image through some paths as an example path from 16 directions toward the target pixel (Hirschmüller, 2008). The energy functional for SGM can be written as

$$S(\mathbf{p}, d) = \sum_{r} L_r(\mathbf{p}, d) \tag{1}$$

where the costs L_r including two penalties for the nearest neighbours and farthest ones computed from different predefined paths toward target pixel ($\mathbf{p}$) are aggregated to produce a global energy. Disparity d is estimated when the global energy becomes minimum through hierarchical matching procedure:

$$\min_d \{S(\mathbf{p}, d)\} \tag{2}$$

In most 3D reconstruction pipelines, the first step is to construct an epipolarity condition for the image pairs. The epipolarity is useful to efficiently speed up the matching process and enhance the robustness against incorrect matches (Morgan, 2004). Several studies investigated the epipolarity constraint for optical-optical image pairs and radargrammetric stereo SAR data takes (Oh et al., 2010, Gutjahr et al., 2014).

With establishing the epipolarity constraint, search spaces for finding the correct conjugate pixel are reduced from 2D to 1D and the energy functional can be constructed for pixels only located on the epipolar line in the corresponding image. The epipolarity constraint can always be found for stereo optical images acquired by frame type cameras, but it becomes more complex in case of stereo images captured by a linear scanning pushbroom sensor such as WorldView-2. For SAR-optical image pairs, it is also beneficial to take the advantage of the epipolarity constraint especially for dense matching by SGM. Its applicability to the SAR-optical multi-sensor setup will be discussed in Section 4.

2.2 SAR-Optical Multi-Sensor Block Adjustment

Before carrying out the dense matching process, a multi-sensor block adjustment approach should be implemented to align the optical image to the SAR image. The main objective is to improve the absolute accuracy of the WorldView-2 image exploiting the high localization accuracy of the TerraSAR-X data take. This can be obtained by modification of RPCs delivered with WorldView-2 data through the block adjustment pipeline with RPCs (Grodecki and Dial, 2003, d'Angelo and Reinartz, 2012). The output of the block adjustment will lead to bias compensation of the optical image. Consequently, the positions of the epipolar lines will be shifted to accurate locations.

By the bundle adjustment, the RPCs are modified by an affine model as follows:

$$\begin{aligned} \Delta r &= a_0 + a_1 r + a_2 c \\ \Delta c &= b_0 + b_1 r + b_2 c \end{aligned} \tag{3}$$

where r and c are rows and columns of points and Δr and Δc are added to rational functions to modify the location of points.

For implementing block adjustment, at first, RPCs must be generated for the TerraSAR-X image. It can be done by the so-called Terrain-Independent approach (Zhang et al., 2011), in which, RPCs are produced using Virtual Ground Control Points (VGCPs) generated by evaluating the Range-Doppler equations for different height levels. After RPC estimation, tie points are selected between a TerraSAR-X image and a WorldView-2 image. The tie points can either be selected by a sparse key point matching method or manually. At the end, the block adjustment equations can be constituted and solved by least squares.

3. EXPERIMENTS AND RESULTS

3.1 Dataset

In this study, SGM applied to multi-sensor image pairs is evaluated based on a high resolution spotlight TerraSAR-X image and a WorldView-2 image acquired over the city of Munich, Germany. The properties of the two images are mentioned in Tab. 1. After the bundle adjustment, from these images two sub-scenes with size of 1000m $\times$ 1500m are cropped from an overlapped area for dense matching. Figure 1 displays the selected image sub-scenes for 3D reconstruction. A nearly zero off-nadir view angle of the WorldView-2 image makes it ideal for stereogrammetry while the TerraSAR-X image was acquired with considerable off-nadir angle because of the SAR-inherent imaging geometry. Both images are resampled to 1m $\times$ 1m pixel size to ease the matching process by enhancing the general image similarity. Moreover, due to different time acquisitions between optical and SAR datasets some changes between two images are expected and it makes the matching process unsuccessful in areas with changes.

3.2 SGM Cost Functions

As explained in Section 2.1, SGM uses a cost function to find the disparity map. Usually, signal-based cost functions such as normalized Cross Correlation (NCC), Mutual Information (MI) etc. are preferred because of their low computational costs in comparison to descriptor-based similarity measures. Among signal-based cost functions, MI is often used for images with complex illumination relationship such as SAR-optical image pairs (Viola and Wells III, 1997). It is formed based on the entropies of the sources images by

$$MI = H_i + H_j - H_{i,j} \tag{4}$$

where $H_k (k = i, j)$ are the entropies of the source images and $H_{i,j}$ is the joint entropy of the two images. Matched points are points with higher MI information that can be obtained through minimizing the joint entropy.

Another similarity measure that will be used in the heart of SGM as cost function is Census that actually acts as non-parametric transformation (Humenberger et al., 2010). The Census similarity measure for image $\mathbf{I}$ and $\mathbf{I}'$ can be defined as:

$$T[I, I'] = \bigotimes_{i=-p}^{p} \bigotimes_{j=-q}^{q} \xi(I(u, v), I'(u', v')) \tag{5}$$

where i and j are indices of pixels belonging to a window with the center of (u, v) for the left images and (u', v') for the right image, $I(u, v)$ and $I'(u', v')$ give gray values in considered locations and $\otimes$ is a bit concatenation operator. The ξ is defined as:

$$\xi(x, y) = \begin{cases} 0 & \text{if } x \leq y. \\ 1 & \text{if } x > y. \end{cases} \tag{6}$$

Furthermore, the weighted sum of both MI and Census are useful for 3D reconstruction in urban areas, especially for a sharper appearance of the reconstructed building outlines. The weighted similarity measure can be defined as (Zhu et al., 2011)

$$SM = \alpha\, MI + (1 - \alpha)\, Census \tag{7}$$

Area	Sensor	Acquisition Mode	Off-Nadir Angle	Ground Pixel Spacing (m)	Acquisition date
Munich	TerraSAR-X	Spotlight	22.99	0.85 × 0.45	03.2015
	WorldView-2	Panchromatic	5.2	0.5 × 0.5	07.2010

Table 1. Specifications of the TerraSAR-X and WorldView-2 images used in this study for dense matching

Figure 1. The overlapped study patches selected from the WorldView-2 image (shown in the left) and the TerraSAR-X image (shown in the right) from images acquired over Munich city

where α changes from 0 to 1 to weigh the effect of Census cost in relation to MI.

Each aforementioned cost function can be employed in SGM to find the optimum disparity map at the end. The conjugate pixels' locations are computed from the estimated disparity map with the minimum aggregated cost. The output will be a disparity map in the reference sensor frame usually considered as left image in the stereogrammetric 3D reconstruction procedure.

3.3 SGM Results

The output of the SAR-optical SGM is a disparity map in the reference sensor geometry, which is the WorldView-2 image in our case. This disparity map is then transformed to a world coordinate system such as UTM. For sake of comparison, results for all those cost functions described in Section 3.2 are created. In all cases, the 3D reconstruction of TerraSAR-X and WorldView-2 images finally produced a sparse rather than dense point cloud over urban areas. Figure 3 shows the achieved point cloud from TerraSAR-X/WorldView-2 image pairs over the city of Munich. Quality assessment is performed respective to a reference LiDAR point cloud with a density of one point per square meters.

Different methods were employed for comparing the generated point cloud and the reference one such as nearest neighbor distance, Least Square (LS) plane fitting and triangulation approaches.

Table 2 provides the height accuracy of the produced point cloud through the stereogrammetric 3D reconstruction by SGM-MI dense matching of TerraSAR-X and WorldView-2 image pairs with different metrics: Mean, Standard Deviation (STD) and RMSE.

Figure 2 shows the performance of using both MI and Census in SGM. The height RMSE of the achieved point cloud for each weight rate was calculated based on the LS-plane fitting method and outliers were detected according to the LE90 criterion. The position of reconstructed points with accuracy better than one pixel (1m) in respective to a reference LiDAR data is displayed in Fig. 4

Area: Munich	Mean	STD	RMSE
Nearest Neighbour	0.196	2.964	2.970
LS-Plane Fitting	0.08	2.652	2.653
Triangulation	0.173	2.977	2.982

Table 2. The height accuracy assessment of the achieved point clouds from the 3D reconstruction of optical-SAR image pair over study subsets in Munich by different methods

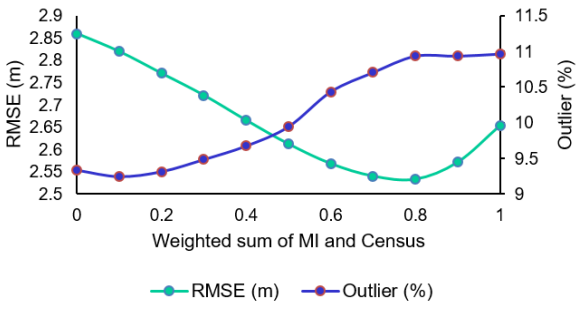

Figure 2. The Height residual (im meter) of the point cloud generated by SGM and the weighted sum cost function.

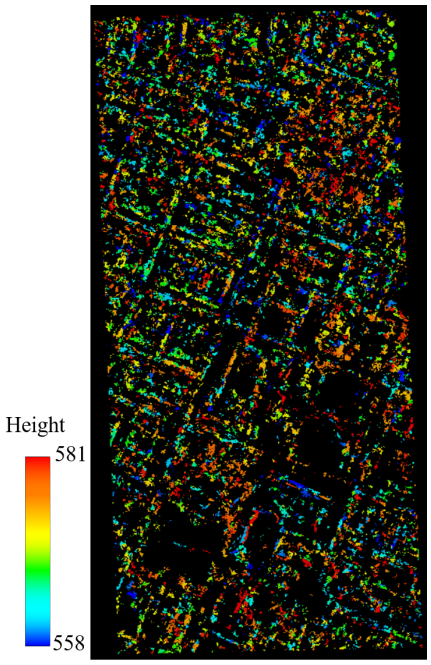

Figure 3. The point cloud reconstructed from TerraSAR-X/WorldView-2 images over the Munich city.

4. DISCUSSION

In order to assess the feasibility of SAR-optical stereo for WorldView-2 and TerraSAR-X data of urban scenes in general, and the applicability of SGM for multi-sensor matching in particular, several points need to be taken into account. These are discussed in the following.

4.1 Validity of the Epipolarity Constraint

As a necessary pre-requesite for multi-sensor matching, we investigated the epipolarity constraint for SAR-optical image pairs such as TerraSAR-X and WorldView-2 images. At first, one specific point (e.g. a point located in the center of the image or in the corner of the building of the Munich central train station in the TerraSAR-X image) is selected. Then the epipolar line corresponding to this point is achieved on the WorldView-2 image using the rational functions and a height step of e.g. 10 m. For checking the epipolar conjugacy from the obtained epipolar line on WorldView-2, two arbitrary points are selected and the corresponding epipolar lines are constructed on TerraSAR-X images. Ideally, two new epipolar lines should be coincident. While the ideal epipolarity situation occurs in frame type images, for linear scanning sensors such as WorldView-2 as well as for SAR sensors, the epipolarity is not ideally established. However, the analysis on the TerraSAR-X and WorldView-2 image pair identifies that the epipolarity can be assumed with sub-pixel accuracy. As a result, the epipolarity constraint can be employed for dense matching of the stereo image pair. Figure 5 displays epipolar lines constructed by RPCs for both, TerraSAR-X image and WorldView-2 images.

4.2 Use of Block Adjustment

As illustrated in Section 4.1, the epipolarity condition can be established for TerraSAR-X and WorldView-2 image pairs. However, the location of epipolar lines in the WorldView-2 image can be modified to a more exact position. This can be performed by

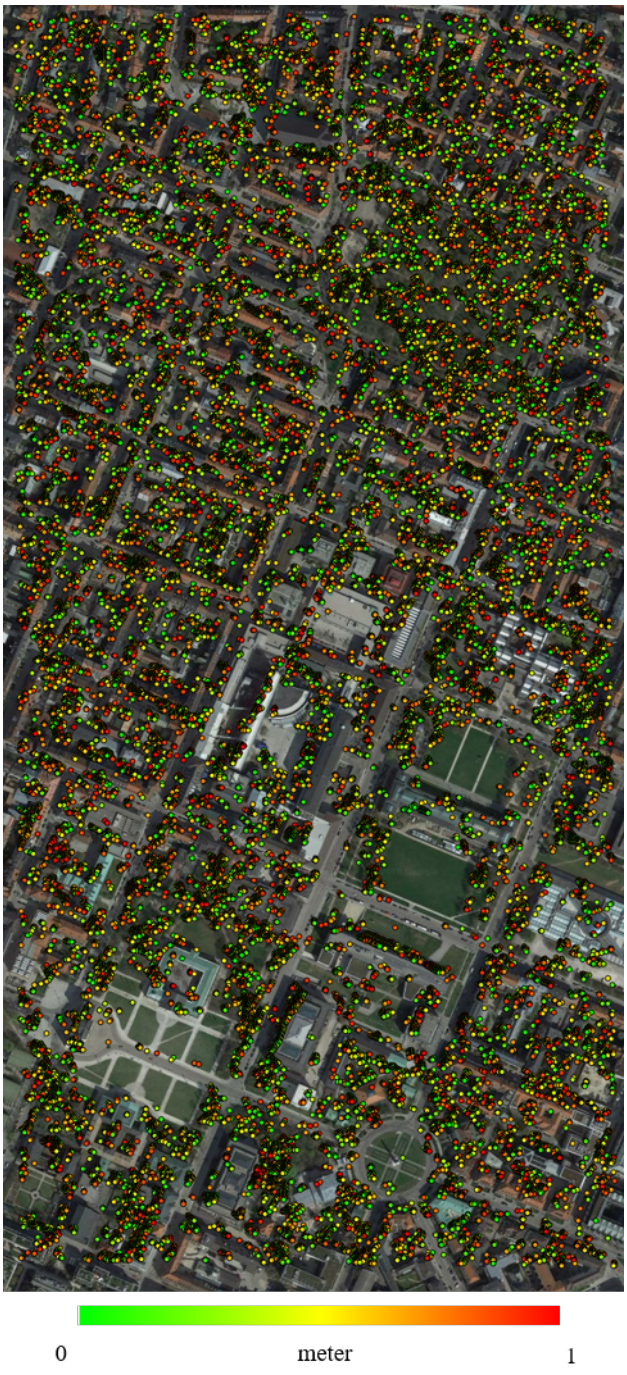

Figure 4. The position of points with accuracy lower than one pixel (1m) generated by SGM and MI cost function over Munich city with respect to a LiDAR reference point cloud. The SGM output is a sparse point cloud with an average density of about 2 points per 10 square meters and an accuracy of better than sub pixel for 30% of the points and finer than 3m for 70% of the points.

the block adjustment with only two bias terms (as shifts in the sample and line directions) that displace the epipolar line to correct position respective to the TerraSAR-X imaging accuracy.

The block adjustment process with RPCs needs some conjugate points between two reference (TerraSAR-X) and target (WorldView-2) images as tie points. Consequently, eight tie points were se-

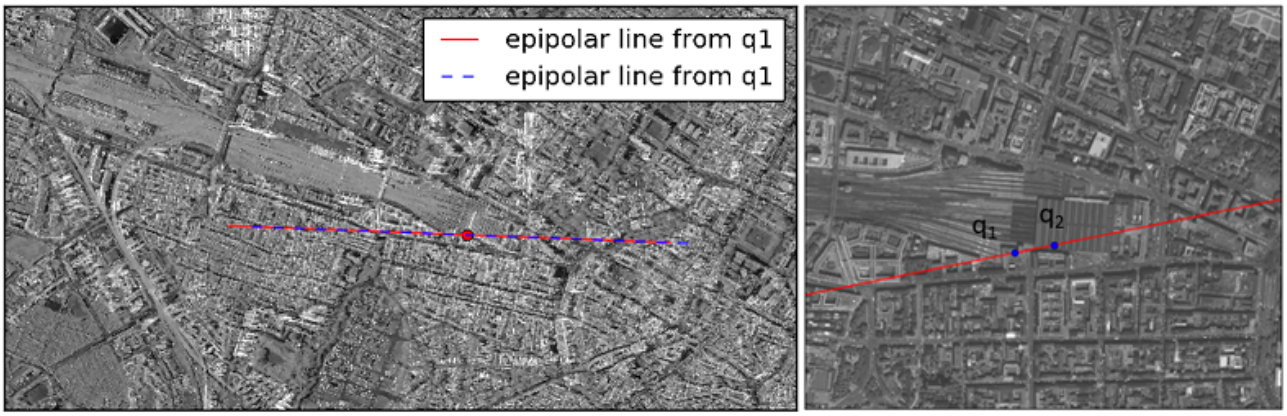

Figure 5. Corresponding Epipolar lines on the TerraSAR-X image (in the left) derived from two points; q1 and q2 on the epipolar line of the WorldView-2 image (in the right)

lected manually. Table 3 represents the terms calculated for coregistration of the WorldView-2 image to the TerraSAR-X image that identifies the major bias is in row direction. Figure 6 illustrates after the block adjustment and coregistration, the epipolar line shifts to new location corresponding to the bias calculated for WorldView-2 RPCs.

4.3 Efficiency of Different Similarity Measures

Different similarity measures like, MI, Census, and weighted sum of MI and Census were examined. The results demonstrate that MI is more stable than Census. The height RMSE decreases using weighted sum especially by increasing the weight of Census but the number of outliers is going up. On the other hand finding the weight value that trades off between the number of outliers and the final accuracy is not easily possible. In normal cases of 3D reconstruction of optical stereo image pairs, the optimal weight can be estimated by 3D visualization and inspection over the footprints of buildings while in the SAR-optical case there is no perfect disparity map to visualize. However, using Census can produce a point cloud with more outliers but less noise in comparison to MI. In conclusion, a similarity measure specifically designed for SAR-optical matching is needed.

4.4 Applicability of SGM for SAR-optical Stereogrammetry

The results show the possibility of 3D reconstruction by semi-global matching of a TerraSAR-X and WorldView-2 image pair. The output is a sparse point cloud with an average density of about 2 points per 10 square meters and an accuracy of better than 3m for 70% of the points. In spite of these numbers, the final generated point cloud still contains a large amount of noisy points, and there is no clear trend regarding the locations of well reconstructed points (cf. Fig. 4). On the other hand, Figure 3 shows that the spatial locations of the reconstructed points do contain information about the scene structure.

An important factor for the comparably low matching success rate and the noisy result is the difference in imaging geometries for SAR and optical sensors (Qiu et al., 2018), which causes many points not to be sensed by the optical sensor while they are perfectly observed by the side-looking SAR sensor. This holds especially for points located on the facades of buildings. Due to these severe differences in imaging geometries causing very different appearances for elevated objects in both images, the isotropic search strategy at the core of SGM is unfavorable and will often lead to false matches even for pixels with low cost. Thus, a crucial necessity in SAR-optical stereogrammetry will be the development of a search strategy, which is inspired by SGM, but takes the mentioned multi-sensor peculiarities into account.

5. CONCLUSION

In this paper, the applicability of semi-global matching for 3D reconstruction from TerraSAR-X and WorldView-2 image pairs was investigated. Similar to other stereogrammetric 3D reconstruction cases (optical stereogrammetry and radargrammetry), a framework was designed and implemented for SAR-optical multi-sensor stereogrammetry. At first, RPCs were estimated for the TerraSAR-X image using the range-Doppler equations. It eases the further processing steps, such as estabilishing the epipolarity constraint and block adjustment. The investigation illustrated that the epipolarity constraint existed for the considered image pair subset. Finally, SGM with different similarity measures was carried out to achieve point clouds of rather sparse and noisy nature.

In order to make 3D reconstruction of urban areas using SAR-optical stereogrammetry produce better results in the future, both new similarity measures tailored to multi-sensor image comparison, as well as adapted matching strategies taking multi-sensor geometry differences into account are needed.

ACKNOWLEDGEMENTS

The authors would like to thank the Bavarian Surveying Administration for providing the LiDAR data.

REFERENCES

d’Angelo, P. and Reinartz, P., 2012. DSM based orientation of large stereo satellite image blocks. In: N. El-Sheimy and M. Shortis (eds), XXII ISPRS Congress 2012, ISPRS Archive, Vol. XXXIX-B1, pp. 209–214.

DigitalGlobe, 2018. Accuracy of WorldView products. `https://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/38/DG_ACCURACY_WP_V3.pdf`. (Accessed 03.18).

Eldhuset, K. and Weydahl, D. J., 2011. Geolocation and stereo height estimation using TerraSAR-X spotlight image data. IEEE Transactions on Geoscience and Remote Sensing 49(10), pp. 3574–3581.

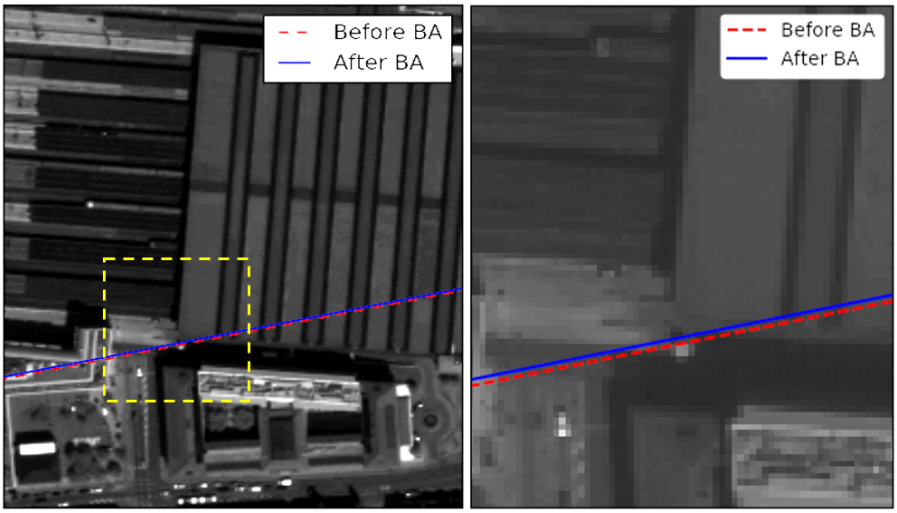

Figure 6. The epipolar line location before (dashed red) and after (blue) the block adjustment on WorldView-2 image

Area	Bias coefficients		STD		No. of Tie Points
	Row	Column	Row	Column	
Munich	-2.47	-0.53	4.96	6.76	8

Table 3. The block adjustment results

Grodecki, J. and Dial, G., 2003. Block adjustment of high-resolution satellite images described by rational polynomials. Photogrammetric Engineering & Remote Sensing 69(1), pp. 59–68.

Gutjahr, K., Perko, R., Raggam, H. and Schardt, M., 2014. The epipolarity constraint in stereo-radargrammetric DEM generation. IEEE Transactions on Geoscience and Remote Sensing 52(8), pp. 5014–5022.

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(2), pp. 328–341.

Humenberger, M., Engelke, T. and Kubinger, W., 2010. A Census-based stereo vision algorithm using modified semi-global matching and plane fitting to improve matching quality. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, pp. 77–84.

Morgan, M. F., 2004. Epipolar resampling of linear array scanner scenes. University of Calgary, Department of Geomatics Engineering.

Oh, J., Lee, W. H., Toth, C. K., Grejner-Brzezinska, D. A. and Lee, C., 2010. A piecewise approach to epipolar resampling of pushbroom satellite images based on RPC. Photogrammetric Engineering & Remote Sensing 76(12), pp. 1353–1363.

Qiu, C., Schmitt, M. and Zhu, X. X., 2018. Towards automatic SAR-optical stereogrammetry over urban areas using very high resolution imagery. ISPRS Journal of Photogrammetry and Remote Sensing 138, pp. 218 – 231.

Schmitt, M. and Zhu, X. X., 2016. Data fusion and remote sensing: An ever-growing relationship. IEEE Geoscience and Remote Sensing Magazine 4(4), pp. 6–23.

Schmitt, M., Tupin, F. and Zhu, X. X., 2017. Fusion of SAR and optical remote sensing data - challenges and recent trends. In: IGARSS 2017.

Viola, P. and Wells III, W. M., 1997. Alignment by maximization of mutual information. International Journal of Computer Vision 24(2), pp. 137–154.

Zhang, L., He, X., Balz, T., Wei, X. and Liao, M., 2011. Rational function modeling for spaceborne SAR datasets. ISPRS Journal of Photogrammetry and Remote Sensing 66(1), pp. 133 – 145.

Zhu, K., d’Angelo, P. and Butenuth, M., 2011. A performance study on different stereo matching costs using airborne image sequences and satellite images. In: U. Stilla, F. Rottensteiner, H. Mayer, B. Jutzi and M. Butenuth (eds), Photogrammetric Image Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 159–170.