

# Towards the Reconstruction of Prismatic Building Models by SAR-Optical Stereogrammetry

Hossein Bagheri<sup>1</sup>, Michael Schmitt<sup>1</sup>, Xiao Xiang Zhu<sup>1,2</sup>

<sup>1</sup>Signal Processing in Earth Observation, Technical University of Munich (TUM)  
Arcisstr. 21, 80333 Munich, Germany

<sup>2</sup>Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR)  
Münchener Str. 20, 82234 Wessling, Germany

**Abstract**—A huge archive of very high-resolution SAR and optical satellite imagery acquired by different remote sensing satellites provides the opportunity to explore the possibility of 3D-reconstruction by multi-sensor stereogrammetry. This paper investigates the potential of SAR-optical stereogrammetry over urban areas using very-high-resolution imagery acquired by TerraSAR-X and Worldview-2. Furthermore, the potentials and challenges of deriving simple prismatic building models by combining the stereogrammetry results and OpenStreetMap building footprints are discussed. The results of this data fusion research demonstrate the possibility of using SAR-optical stereogrammetry for urban 3D reconstruction at level-of-detail 1.

**Index Terms**—data fusion, stereogrammetry, 3D reconstruction, dense matching, urban areas, city models

## I. INTRODUCTION

One particular interest in remote sensing is the 3D reconstruction of urban areas for diverse applications such as 3D city modelling, urban and crisis management etc. A typical method for reconstructing urban areas on a large scale is to employ stereo optical imagery provided by high-resolution space-borne sensors in an ideal acquisition situation. However, because of limitations in acquiring those ideal images such as cloud effects as well as limited absolute localization accuracy, optical stereo might not always be the optimal choice. In contrast, the amplitude images provided by SAR sensors do not suffer from the aforementioned issues, and can thus provide input to time-critical 3D reconstruction tasks. Regarding the growing archive of very high-resolution SAR and optical imagery, developing a framework that takes advantage of both SAR and optical imagery can provide a great opportunity to produce 3D spatial information over urban areas as an application of data fusion in remote sensing [1].

In the literature, a few papers can be found that dealt with the combination of SAR and optical imagery for the 3D reconstruction of urban objects, e.g. [2]. However, none of them has exploited multi-sensor stereogrammetry for this purpose yet. In this research, we focus on the potential of 3D reconstruction from very high-resolution SAR-optical image pairs such as TerraSAR-X/WorldView-2 through a dense matching process. To the stereogrammetric processing chain, a simple approach for the generation of 3D prismatic building models by combining the derived height information with OpenStreetMap building footprints is added.

## II. RECONSTRUCTING BUILDING MODELS FROM SAR-OPTICAL STEREOGRAMMETRY

A full framework for stereogrammetric 3D reconstruction from SAR-optical image pairs was presented in [3], [4]. It consists of several steps: generating rational polynomial coefficients (RPCs) for each image to replace the different physical imaging models by a homogenized mathematical model; RPC-based multi-sensor block adjustment to enhance the relative orientation between both images; establishing a multi-sensor epipolarity constraint to reduce the matching search space from 2D to 1D.

The main step in SAR-optical stereogrammetry, however, is to estimate disparity values by using a dense matching algorithm such as semi-global matching (SGM). In the following, more details of SAR-optical dense matching along with experimental results of TerraSAR-X/WorldView-2 stereogrammetry over urban areas will be presented. After that, the potentials and challenges of using the thus-produced 3D point clouds for the generation of simple prismatic models are discussed.

### A. SAR-Optical Dense Matching

The core challenge in SAR-optical stereogrammetry is to find disparity maps between two images by using a dense matching algorithm. For the presented research, we have investigated the application of classical SGM for that purpose. SGM computes the optimum disparity maps by minimizing an energy functional which is constructed by a data and a fidelity term [5]. While the data term is defined by a similarity measure, the fidelity term employs two penalties to smooth the final disparity map. Because of aggregating cost values computed by a cost function in the heart of SGM along with a regularizing smoothness term, SGM is relatively robust against noise, which seems to make it a good choice for SAR-optical stereogrammetry. However, some specific settings must be considered to adapt SGM to SAR-optical image pairs: One of the main impact factors on the performance of SGM for dense matching of SAR-optical imagery is the robustness and strength of cost functions used in SGM to measure the similarity between SAR and optical images because of the complicated radiometric relationship. According to [6], pixel-wise Mutual information (MI), and Census are more appropriate for difficult illumination relationships than, e.g.

normalized cross-correlation (NCC). In this research, we have investigated all conventionally used similarity measures, i.e. pixel-wise MI, Census, their weighted sum, and NCC.

Next, the minimum and maximum disparity values are used to limit the length of the search domain along epipolar curves. Those values can be achieved from an external DEM like SRTM or selected according to the mean height of the study area, considering a height threshold suited for the target urban area.

The 3D-reconstruction results achieved with these SGM settings are shown in Section III.

### B. Derivation of Prismatic Building Models

One of the most interesting applications of modern photogrammetric 3D reconstruction is to produce 3D building models [7]. There are different levels of modeling which have been gathered under the standard of the City Geography Markup Language (CityGML). These are summarized in [8].

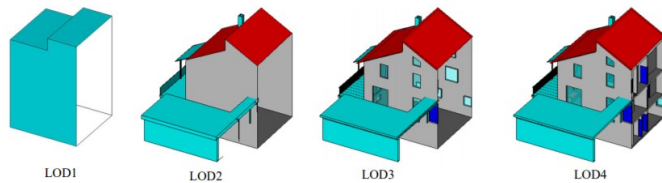


Fig. 1: Different Levels of building models in CityGML [9]

Figure 1 displays different levels-of-detail as defined in the CityGML standard. As shown in this figure, the lowest level of detail (LOD) is 1 (LOD1), which simply models buildings as boxes without considering any details. Thus, for LOD1 only the outlines of buildings along with height information are required. Regarding the output of SAR-optical stereogrammetry being a relatively noisy and sparse point cloud, an inclusion of auxiliary data is advised: We thus chose to extract the building footprints from the OpenStreetMap building layer, while the height information can be derived from the output of SAR-optical stereogrammetry. This is done by assigning the median height of all points found within a building outline to the footprint and thus extruding the footprint into the height axis as well.

## III. EXPERIMENTS AND RESULTS

### A. Test Data

For the experiments presented in this paper, a high-resolution TerraSAR-X/WorldView-2 image pair, acquired over the city of Berlin, Germany, is used. For the pre-processing, first, the SAR image was filtered by a non-local filter to reduce the speckle. After that, they were resampled to  $1\text{m} \times 1\text{m}$  pixel size to homogenize the study scenes with respect to better similarity estimation. After multi-sensor bundle adjustment, sub-images from the overlapped part of the study area were selected. These sub-images are displayed in Fig. 2. The specifications of the TerraSAR-X and WorldView-2 images are provided in Tab I.



Fig. 2: Display of study subsets located in Berlin, left: WorldView-2, right:TerraSAR-X

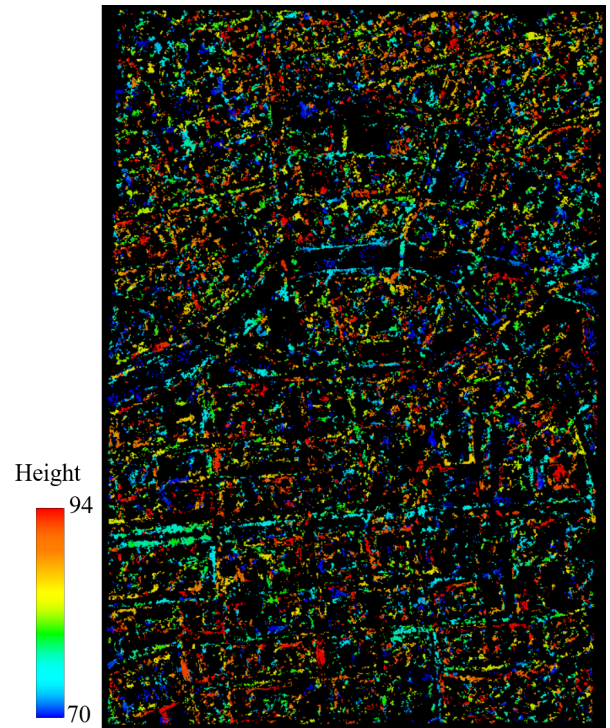


Fig. 3: Achieved point cloud from stereogrammetric 3D reconstruction of TerraSAR-X/WorldView-2 over the Berlin study area

### B. Dense Matching Results

By using SGM for dense matching of the TerraSAR-X and WorldView-2 image pair, a sparse point cloud could be produced. As mentioned in Section II-B, we tried different similarity measures in the SGM data term. As an example, the achieved point cloud by using MI as a similarity measure is shown in Fig. 3.

To validate the accuracy of the resulting 3D point clouds, we employed an accurate LiDAR point cloud with an average density of 6 points per square meters and a vertical accuracy of 20cm. For accuracy calculation, after Least Square (LS) plane fitting on  $k$  (here:  $k = 6$  points) nearest neighbors of

TABLE I: Specifications of the TerraSAR-X and WorldView-2 images

Sensor	Acquisition Mode	Off-Nadir Angle (°)	Ground Pixel Spacing (m)	Acquisition date
TerraSAR-X	Spotlight	36.11	$0.17 \times 0.45$	04.2016
WorldView-2	Panchromatic	29.1	$0.5 \times 0.5$	05.2013

each target point in the reference point cloud, the Euclidean distance between the target point to the fitted reference plane was measured along different directions. Tab. II summarizes accuracy assessments of the reconstructed point clouds using different similarity measures along different coordinate axes by LS plane fitting. Additionally, the mean absolute difference between the achieved point cloud respective to the LiDAR data is applied for total accuracy evaluation. As represented in Tab. II, the MI similarity measure gives the best results which means that pixel-wise MI has better performance than other classical similarity measures for complicated matching scenarios such as SAR-optical image pairs. Still, the achieved 3D accuracy is slightly worse than the pre-processed pixel size, thus indicating the need for more sophisticated similarity measures, which are specifically dedicated to the SAR-optical multi-sensor case. Such a similarity measure could, e.g., result from deep learning [10].

To reduce the influence of outliers, we applied some post-processing to the reconstructed point clouds, exploiting an external DEM such as SRTM to detect inaccurate points. For this, those points, whose heights deviate from SRTM by more than a specified threshold are removed. As our assessments demonstrate, using this outlier removal can significantly improve the quality of the obtained point cloud (cf. Tab II), in spite of the low resolution of the SRTM dataset.

### C. LOD1 Building Model Result

Figure 4 displays an LOD 1 city model consisting of prismatic building models generated by combining the height information derived from SAR-optical stereogrammetry and building footprints provided by OpenStreetMap. The height accuracy of this LOD 1 model was validated by comparing it a similarly derived city model with heights taken from the LiDAR point cloud. The results illustrate that the median and mean absolute height differences between the produced SAR-optical model and the LiDAR city model are 1.56 m and 0.84 m, respectively.

## IV. DISCUSSION

### A. Potential and Limitations of SAR-Optical Dense Stereogrammetry

The achieved results demonstrate the possibility of 3D reconstruction from very high-resolution SAR-optical image pairs by stereogrammetry in urban areas. The best result was obtained by using MI as similarity measure, while using NCC leads to a point cloud with significantly worse accuracy. This is partially due to the fact that area-based similarity measures such as NCC are not really suitable for dense image matching, since the required rectangular window also acts as some kind of smoothing operator. However, as shown in Fig. 3, the point

cloud from pixel-wise MI-based dense matching also suffers from the noise effects. Apart from the difficult geometric relationship between SAR and optical imagery, their different radiometric appearance makes the matching very challenging. Since the applied similarity measures are classic measures that are not specifically suited for the SAR-optical multi-sensor matching case yet, this limits the matching success rate. In addition, it has been shown that only few scene parts are jointly visible to the SAR and the optical sensors at all in typical urban scenarios.

In spite of all this, SAR-optical stereogrammetry led to a point cloud with a median accuracy of about 1.5 to 2m and the 25%-quantile of best points well in the sub-pixel accuracy domain, thus indicating the general feasibility of the multi-sensor stereogrammetry concept.

### B. Potential of Building Model Derivation from SAR-Optical Point Clouds

There are different challenges relevant to 3D building model generation as well as an accuracy evaluation of the produced model which should be discussed. Firstly, the heights located in an outline of a building should be consistent. Some inconsistencies come from noise produced during the SAR-optical dense matching process. Moreover, regarding shapes of roofs particularly inclined types, the location of points differs from the topmost toward the lowest part. Then, making a decision, which cluster of points are suitable for 3D reconstruction in the LOD 1 level is a challenge. In this research, we simply used median as a robust estimator against the large heights changes of reconstructed points located in a footprint of building.

For validating accuracy of the produced 3D model, we also applied median of absolute height differences in addition to mean absolute height differences which makes the assessment more robust against the large heights differences. In other words, by using median, the heights of partially high-rise buildings are discarded due to domination of noise induced by such as layover and shadow effects. However, employing both metrics, the evaluation results demonstrate the potential of SAR-optical stereogrammetry for LOD1 model generation.

## V. SUMMARY AND CONCLUSION

In this research, the possibility and potential of urban 3D reconstruction from TerraSAR-X and WorldView-2 image pairs through a stereogrammetry pipeline was investigated. Conventional semi-global matching with different similarity measures to drive the cost function was carried out to achieve point clouds with a median accuracy of about 1.5 to 2m and the 25%-quantile of best points well in the sub-pixel accuracy domain. Then, SRTM data were used to remove outliers from the point clouds and further improving the mean



TABLE II: Accuracy assessment of reconstructed point clouds using different similarity measures with respect to LiDAR reference

Similarity Measures	Mean (m)			STD (m)			RMSE (m)			Mean (m)
	X	Y	Z	X	Y	Z	X	Y	Z	d
MI	<b>0.00</b>	-0.04	<b>0.27</b>	<b>1.57</b>	<b>1.69</b>	<b>3.09</b>	<b>1.57</b>	<b>1.69</b>	<b>3.10</b>	<b>2.75</b>
Census	0.01	-0.01	0.66	1.64	1.86	3.33	1.64	1.86	3.40	2.94
NCC	-0.01	-0.02	0.70	1.76	2.04	3.66	1.76	2.04	3.73	3.26
MI+Census	-0.01	<b>0.00</b>	0.61	1.62	1.78	3.25	1.62	1.78	3.31	2.86

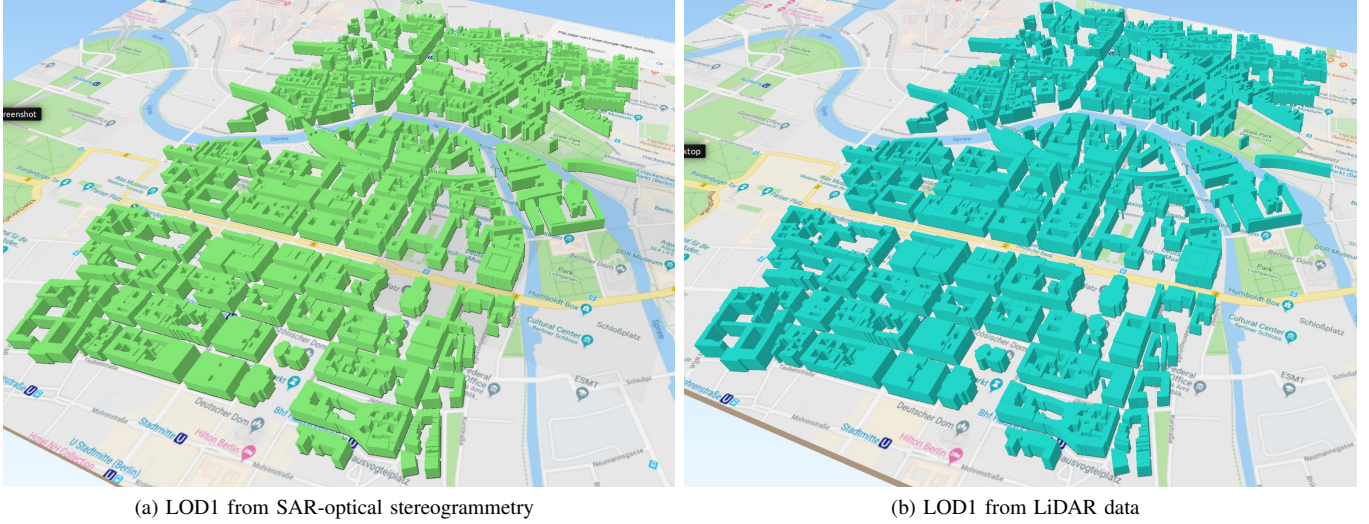


Fig. 4: 3D prismatic model generation with an integration of height information derived from SAR-optical stereogrammetry (a) and building footprint layer from OpenStreetMap, (b): LOD1 model from LiDAR data

TABLE III: Accuracy assessment of point clouds after SRTM-based outlier removal

Area	Point Cloud	Mean (m)
Berlin	original	2.75
	filtered	2.35
	SRTM	2.65

point cloud accuracy. Finally, simple prismatic 3D building models were generated by combining heights produced by SAR-optical stereogrammetry with footprints of buildings provided by OpenStreetMap. These preliminary results demonstrate the possibility of prismatic building model generation (at LOD 1 level) through SAR-optical stereogrammetry and also a possible way for removing noise from SAR-optical stereogrammetry-derived heights.

#### ACKNOWLEDGMENT

The authors want to thank everyone, who has provided test data for this research: European Space Imaging for the WorldView-2 image, DLR for the TerraSAR-X images, and Land Berlin (EU EFRE project) for the LiDAR reference data of Berlin.

#### REFERENCES

[1] M. Schmitt and X. X. Zhu, "Data fusion and remote sensing: An ever-growing relationship," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, pp. 6–23, Dec 2016.

[2] J. D. Wegner, J. R. Ziehn, and U. Soergel, "Combining high-resolution optical and InSAR features for height estimation of buildings with flat roofs," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 5840–5854, Sept 2014.

[3] H. Bagheri, M. Schmitt, P. d'Angelo, and X. X. Zhu, "A framework for SAR-Optical stereogrammetry over urban areas," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018. in press.

[4] H. , M. Schmitt, P. d'Angelo, and X. X. Zhu, "Exploring the applicability of semi-global matching for SAR-optical stereogrammetry of urban scenes," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2, pp. 43–48, 2018.

[5] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 328–341, Feb 2008.

[6] H. Hirschmüller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 1582–1599, Sept 2009.

[7] H. Arefi, *From LiDAR point clouds to 3D building models*. PhD thesis, Universitätsbibliothek der Universität der Bundeswehr München, 2009.

[8] T. H. Kolbe, G. Gröger, and L. Plümer, *CityGML: Interoperable Access to 3D City Models*, pp. 883–899. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[9] Karlsruhe Institute of Technology (KIT), Institute for Applied Computer Science, "CityGML example FZK-Haus." [http://www.citygmlwiki.org/index.php?title=FZK\\_Haus](http://www.citygmlwiki.org/index.php?title=FZK_Haus), 2018. (Accessed 10.18).

[10] L. H. Hughes, M. Schmitt, L. Mou, Y. Wang, and X. X. Zhu, "Identifying corresponding patches in SAR and optical images with a Pseudo-Siamese CNN," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, pp. 784–788, May 2018.