Ingenieurfakultät Bau Geo Umwelt
Technische Universität München

Fusion of Multi-sensor-derived Data for the 3D Reconstruction of Urban Scenes

Hossein Bagheri

Vollständiger Abdruck der von der Ingenieurfakultät Bau Geo Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr.-Ing. habil. Richard Bamler

Prüfende der Dissertation:

1. Prof. Dr.-Ing. habil. Xiaoxiang Zhu
2. Priv.-Doz. Dr.-Ing. habil. Michael Schmitt
3. Prof. Dr.-Ing. Peter Reinartz

Die Dissertation wurde am 12.04.2019 bei der Technischen Universität München eingereicht und durch die Ingenieurfakultät Bau Geo Umwelt am 19.08.2019 angenommen.

Abstract

Urban Three Dimensional (3D) reconstruction is one of the favorite remote sensing tasks for different applications. In this regard, a particular interest lies in the generation of a building model with a potential of large-scale coverage. One of the significant remote sensing sources for this purpose is the medium-resolution Digital Elevation Model (DEM). For example, launching a new mission called TanDEM-X has provided a global covering DEM with an unprecedented relative accuracy, which can potentially be applied to 3D building modeling. However, visual and quality inspections reveal that the TanDEM-X DEM quality drops in urban areas because of an inherent imaging property of the Synthetic Aperture Radar (SAR) sensor. In this dissertation, two solutions are proposed to improve the quality of medium-resolution DEMs, such as TanDEM-X DEM in urban areas. The first solution is to fuse DEM with elevations produced by another type of sensor with different properties. Then, the DEM quality can be improved by taking advantage of multi-sensor DEM fusion to integrate instructive properties of input DEMs and reduce the effects of their defects. For instance, Cartosat-1 DEM is an appropriate choice for fusion with TanDEM-X DEM. This multi-sensor DEM fusion is performed by implementing a sophisticated Artificial Neural Network (ANN)-based fusion framework. It consists of three main steps: spatial feature extraction and corresponding height residual estimation respective to LiDAR ground truth data, data preprocessing to generate primary feature-error patterns, and finally inputting those patterns into fully connected ANNs to explore appropriate weight maps corresponding to each input DEM. The results demonstrate the efficiency of the designed fusion framework for improving the quality of TanDEM-X DEM, as well as increased absolute accuracy of Cartosat- 1 DEM. The next potential solution is to fuse multi-modal DEM acquisitions using advanced data fusion techniques instead of using simple Weighted Averaging (WA). For example, WA is currently used in the process of TanDEM-X raw DEM mosaicking for global DEM generation. The idea is to take advantage of L_1 norm Total Variational (TV-L_1) and Huber models for DEM fusion. These models can efficiently be used in urban areas by smoothing noise influences while preserving edges such as building footprints which are frequently found in these areas. The final obtained fused DEM illustrates an excellent performance of variational models for the DEM quality enhancement in urban areas.

Apart from DEMs, another potential remote sensing resource for urban 3D reconstruction on a large scale is spatial information produced from SAR-optical imagery such as TerraSAR-X and WorldView-2. In this dissertation, a 3D reconstruction stereogrammetric framework is developed for this task. This framework includes several steps: generating Rational Polynomial Coefficient (RPC) for SAR imagery, establishing an epipolarity constraint between SAR and optical imagery, developing a multi-sensor block adjustment, generating disparity map by a dense matching algorithm and finally a forward intersection for producing a point cloud. The final results demonstrate the potential of 3D reconstruction from SAR-optical imagery

using the proposed stereogrammetry framework.

Finally, the possibility of generating 3D building model at the first Level Of Details (LOD1) using elevations derived from either enhanced fused DEMs or a point cloud produced by SAR-optical stereogrammetry in combination with building footprints from OpenStreetMap (OSM) is investigated. The results confirm the potential of LOD1 building model generation from those multi-sensor-derived heights and OSM building footprints.

Zusammenfassung

Die städtische dreidimensionale Rekonstruktion ist eine der beliebtesten Fernerkundungsaufgaben für verschiedene Anwendungen. In diesem Zusammenhang liegt ein besonderes Interesse auf der Erstellung von Gebäude Modellen mit einem Potenzial für eine großflächige Abdeckung. Eine der wichtigsten Quellen für Daten für dieses Ziel ist das mittelauflösende digitale Höhenmodell. So hat beispielsweise der Start der neuen TanDEM-X Mission einem global abdeckenden DEM-eine beispiellose relative Genauigkeit verliehen, die für die 3D-Gebäudemodellierung potenziell genutzt werden kann. Visuelle und qualitätive kontrollen zeigen jedoch, dass die TanDEM-X DEM-Qualität in städtischen Gebieten aufgrund der inhärenten Abbildungseigenschaften des SAR-Sensors sinkt. In dieser Dissertation werden zwei Lösungen vorgeschlagen, um die Qualität von mittelauflösenden DEMs, wie dem TanDEM-X DEM, in städtischen Gebieten zu verbessern. Die erste Idee ist, das DEM mit Höhen zu verschmelzen, die von einem anderen Sensortyp mit unterschiedlichen Eigenschaften erzeugt werden. Anschließend kann die Verbesserung der DEM-Qualität durch den Einsatz der Multisensor-DEM-Fusion realisiert werden, um die instruktiven Eigenschaften der DEMs zu integrieren und die Auswirkungen ihrer Fehler zu reduzieren. So ist beispielsweise der Cartosat-1 DEM eine geeignete Wahl für die Fusion mit dem TanDEM-X DEM. Diese Multisensor-DEM-Fusion basiert auf der Implementierung eines anspruchsvollen Frameworks mit ANNs. Diese DEM-Fusion besteht aus drei Hauptschritten: Räumliche Merkmalsextraktion und entsprechende Höhenrestschätzung gemäß den LiDAR-Bodenwahrheitsdaten, Datenvorverarbeitung zur Erzeugung primärer Merkmalsfehlermuster und schließlich Eingabe dieser Muster in vollständig verknüpfte ANNs, um geeignete Gewichtskarten gemäß jeder eingegebenen DEM zu untersuchen. Die Ergebnisse zeigen die Effizienz des Fusionsrahmens zur Verbesserung der Qualität des TanDEM-X DEM und zur Erhöhung der absoluten Genauigkeit des Cartosat-1 DEM. Eine alternative Lösung besteht darin, multimodale DEM-Akquisitionen mit fortschrittlichen Datenfusionstechniken zu fusionieren, anstatt einfache WA zu verwenden. So wird beispielsweise die WA derzeit im Prozess des TanDEM-X Roh-DEM-Mosaiks für die globale DEM-Generierung verwendet. Die Idee ist, die Vorteile der TV-L1 und Huber Modelle für die DEM-Fusion zu nutzen. Diese Modelle können in urban Räume effizient eingesetzt werden, indem Lärmeffekte unter Beibehaltung der in diesen Gebieten üblichen Kanten, wie z.B. Gebäudeaufstandsflächen, geglättet werden. Das final fusionierte DEM veranschaulicht die hervorragende Leistung von Variational modellen zur Verbesserung der DEM-Qualität in städtischen Gebieten.

Neben DEMs ist eine weitere potenzielle Fernerkundungsressource für die großflächige städtische 3D-Rekonstruktion räumliche Informationen, die aus SAR-optischen Bildern, wie TerraSARX und WorldView-2 gewonnen werden. In dieser Arbeit wird ein stereogrammetrisches 3D-Rekonstruktionsgerüst für diese Aufgabe entwickelt. Dieses Framework umfasst mehrere Schritte: Rationelle Polynomkoeffizientengenerierung für SAR-Bilder,

Etablierung einer Epipolaritätsbeschränkung zwischen SAR und optischer Bildgebung, Entwicklung einer Multisensor-Blockanpassung, Disparitätskartengenerierung durch einen dichten Matching-Algorithmus und schließlich ein Vorwärtsschnitt zur Erzeugung einer Punktwolke. Die Endergebnisse zeigen das Potenzial der 3D-Rekonstruktion aus SAR-optischen Bildern unter Verwendung des vorgeschlagenen Stereogrammetrie-Rahmens.

Schließlich wird die Möglichkeit der 3D-Gebäudemodellgenerierung auf der ersten Detailebene unter Verwendung von Elevationen untersucht, die entweder aus verbesserten fusionierten DEMs oder einer Punktwolke stammen, die durch SAR-optische Stereogrammetrie in Kombination mit Gebäude-Footprints aus OSM erzeugt wurde. Die Ergebnisse bestätigen das Potenzial der LOD1-Gebäudemodellgeneration aus diesen multisensorisch gewonnenen Höhen und den von OSM bereitgestellten Gebäudegrundrissen.

Table of contents

Acronyms

- **3D** Three Dimensional
- **ACV** Anisotropic Coefficient of Variation
- **CityGML** City Geography Markup Language Model
- **DEM** Digital Elevation Model
- **DLR** German Aerospace Center
- **DMP** DEM Mosaicking Processor
- **DSM** Digital Surface Model
- **DTM** Digital Terrain Model
- **GCP** Ground Control Point
- **HEM** Height Error Map
- **HoA** Height of Ambiguity
- **HRTI** High-Resolution Terrain Information
- **ICP** Iterative Closest Point
- **InSAR** Interferometric SAR
- **ITP** Integrated TanDEM-X Processor
- **LOD** Level OF Details
- **LE** Linear Error
- **MI** Mutual Information
- **NMAD** Normal Median Absolute Deviation
- **OSM** OpenStreetMAP
- **PU** Phase Unwrapping
- **RMSE** Root Mean Square Error
- **ROF** Rudin, Osher, and Fatemi

- **RPC** Rational Polynomial Coefficient
- **SAR** Synthetic Aperture Radar
- **SGM** Semi-global Matching
- **SRF** Surface Roughness Factor
- **SSE** Sum of Squared Errors
- **STD** Standard Deviation
- **TGV** Total Generalized Variation
- **TPI** Topographic Position Index
- **TRI** Topographic Ruggedness Index
- **TV** Total Variation
- **VGCP** Virtual Ground Control Point
- **VGI** Volunteered Geographic Information
- **VHR** Very High-Resolution
- **WA** Weighted Averaging

1 Introduction

1.1 Motivation

3D reconstruction from remote sensing data has a range of applications across various fields, such as 3D city modeling, urban and crisis management, environmental studies, and geographic information systems. Manifold high-resolution sensors in space provide the possibility of reconstructing natural and human-made landscapes over large-scale areas.

Buildings are one of the main categories of objects in urban scenes, which are modeled for diverse applications such as air pollution simulation, energy consumption estimation, urban heat island detection and many others [1]. Buildings can be modeled in different level of detail gathered under the standard of the City Geography Markup Language (CityGML) [2]. Figure 1.1 schematically displays the amount of details required to be presented at each level.

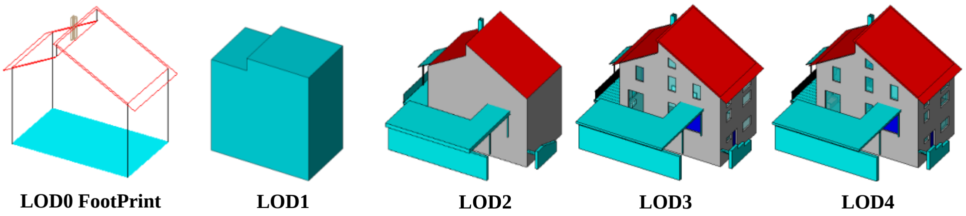


Figure 1.1: Different level of detail of building models according to CityGML 2.0. Image courtesy of [3]

As displayed in Figure 1.1, the first level is LOD1 which describes building models as block models with flat roof structure and provides the coarsest volumetric representation of buildings [4]. Thus, for LOD1 only the outlines of buildings along with height information are required. The next level is LOD2 that represents building shapes with more details. Therefore, LOD2 building reconstruction demands high-resolution data in comparison to the first level. Among the mentioned models, a particular interest lies in generating building models on a large scale at the LOD1 level for diverse applications such as global population density estimation, urban heat island detection, etc. Thus, city modeling with building details at the LOD1 level is quite enough for those applications.

Building height information can be provided using versatile remote sensing data sources such as airborne laser scanning [5], high-resolution optical stereo imagery [6], interferometric DEMs produced from SAR data [7], and others [8]. Conventionally, height information is provided using airborne LiDAR data that leads to highly accurate LOD1 representations of buildings; however, it is computationally expensive to produce wide area covering models. Valuable LiDAR data are usually not available on a large scale. Consequently, heights can come from different sources, and data fusion is an efficient way to exploit complementary advantages and mitigate distinct disadvantages.

One option of data fusion for large-scale height retrieval could be DEM fusion, as there exist DEMs such as TanDEM-X and Cartosat-1, which have different attributes. Moreover, another option would be to generate heights using SAR-optical stereogrammetry.

1.2 Objectives

The main objective of this research is to investigate the potential and possibility of 3D urban reconstruction and subsequently building modeling using remote-sensing-derived geodata. For this purpose, different data fusion techniques are employed to promote the quality of large-scale mapping as well as to produce heights for 3D building modeling of urban areas. In this regard, the secondary objectives of this work can be categorized into two main issues as follows:

- **Large-scale height generation over urban areas through DEM fusion**: Primary produced DEMs called raw DEMs such as TanDEM-X raw DEMs are not usually perfect, especially for large-scale urban mapping, due to existing errors induced from defects of sensors, which consequently influence the final quality of raw products [9]. As a solution, DEM fusion as an application of data fusion in remote sensing can be employed to improve the quality of produced DEMs. For instance, TanDEM-X DEM as a new global DEM with unprecedented accuracy and nearly complete coverage of earth is produced through the TanDEM-X mission using bistatic SAR interferometric data [10]. However, the quality of the achieved TanDEM-X DEM drops in urban areas because of the inherent SAR imaging geometry such as layover and shadowing effects [9].

  One possible solution for improving elevation data is fusion with other available elevation data, which do not suffer from sensor-inherent imaging effects. In the TanDEM-X DEM case, the alternative data are DEMs derived from high-resolution optical stereo imagery such as those acquired by Cartosat-1. Consequently, in order to reach a better result, an efficient multi-sensor-based DEM fusion technique is proposed in this research.

  Another possibility to improve the quality of final DEM is to fuse the acquired multi-modal raw DEMs, i.e.; fusing raw DEMs derived from data-takes with different properties acquired by the same sensor. For instance, the standard TanDEM-X DEM is the output of a processing chain with interferometry, phase unwrapping, data calibration, DEM block adjustment, and raw DEM mosaicking. In the mosaicking step, raw DEMs are fused to reach the target accuracy. A conventional method for doing DEM fusion in DEM Mosaicking Processor (DMP) is WA, in which heights are summed with respect to their weights derived from a Height Error Map (HEM). While the WA approach can realize the predefined goals in DMP for global DEM generation, it does not perform optimally under challenging terrains with complex morphology, such as urban areas, and contains many high-frequency contents such as edges. After WA-based DEM fusion, visualization shows that the outlines of buildings are not perfectly sharp and still some amount of existing noise spoils building footprints. Therefore, this work investigates the application of more sophisticated multi-modal DEM fusion approaches to efficiently preserve the edges and outlines of buildings while removing noise.

- **SAR-optical stereogrammetry for 3D urban reconstruction:** Regarding the growing archive of Very High-Resolution (VHR) SAR and optical imagery, developing a framework that takes advantages of both SAR and optical imagery can provide a great opportunity to produce 3D spatial information over urban areas. This dissertation also focuses on the potential of 3D building reconstruction from VHR SAR-optical image pairs such as TerraSAR-X/WorldView-2 through a dense matching process as a form of cooperative data fusion. In this context, the main idea is to investigate the applicability of the Semi-Global Matching (SGM) algorithm for SAR-optical stereogrammetry and design a framework for accomplishing this task.

Finally, this research investigates the possibility of generating LOD1 building models from both Volunteered Geographic Information (VGI) and remote sensing-derived geodata, which is available on a wide scale. More specifically, , the study exploit building footprints provided by OSM and height data derived from the fusion of different kinds of multi-sensor data.

1.3 Thesis Structure

The main contributions of this cumulative dissertation come from four peer-reviewed journal papers published by the first author. These publications can be found in the appendices.

The dissertation is organized into four chapters. The motivations and objectives of the thesis are already addressed in this chapter. In the following, comprehensive reviews of previous studies corresponding to each of the mentioned objectives are reported in Section 1.4. Chapter 2 reviews some basic concepts and fundamentals required for research implementations. Chapter 3 provides a summary of the contributions of the author. Finally, the conclusions of the implemented investigations and the achievements are discussed in Chapter 4.

1.4 State-of-the-Art

In line with the objectives of the dissertation, a literature review of previous studies is reported in this section within three main categories including previous research on 3D building model generation from large-scale remote sensing-derived geodata, state-of-the-art DEM fusion techniques, and finally, investigations on 3D reconstruction from SAR-optical imagery. Reviews of each category are presented in a distinct section in the following.

1.4.1 3D Building Models from Remote Sensing-derived Geodata

One of the significant requirements for LOD1 building modeling is height information that can be provided by versatile remote sensing data sources such as airborne laser scanning [5], high-resolution optical stereo imagery [6], and interferometric DEMs produced from SAR data [7]. Among different levels for CityGML-based modeling, a particular interest lies in generating building models on a large scale at the LOD1 level. While height information provided by airborne LiDAR data leads to highly accurate LOD1 representations of buildings, high point densities demand high computational loads and expenditure for a large scale modeling. In addition, valuable LiDAR data are usually not available on a large scale. In the literature, few investigations illustrate the possibility of using medium-resolution, large scale covering remote sensing data types for 3D building reconstruction. As an example, the possibility of LOD1-level 3D building model generation from Cartosat-1 and Ikonos DEMs has been investigated in [11]. The researchers used semi and fully automatic methods for building footprint extraction from satellite imagery. Heights were derived from DEMs generated by stereo image pairs. The results demonstrated that the highest accuracy was achieved using the automatic method for building footprint extraction. However, the both methods demanded a long time to perform the process. In another study, Marconcini et al. proposed a method for building height estimation from TanDEM-X data [12]. They implemented an unsupervised approach to discriminate building points from points lying on the ground surface to finally produce a Digital Terrain Model (DTM) of the study area. Then, building heights were estimated by subtracting the produced DTM from the original DEM. Using open DEMs such as SRTM for 3D reconstruction has been evaluated in different studies [13, 14, 15]. It has been concluded that SRTM elevation data could be used for recognizing tall buildings. In a recent investigation, Misra et al. compared different global height data sources such as SRTM, ASTER, AW3D as well as TanDEM-X for digital building height model generation [16].

The investigations mentioned above attempt to extract building footprints directly from DEMs or satellite imagery. Since medium-resolution DEM data are often not detailed and sufficiently accurate to provide sufficient information for modeling individual buildings and also extracting footprints from satellite imagery demands a considerable effort, this work investigates the potential of using footprints provided by OSM for LOD1 building modeling. Furthermore, because of defects and limitations alongside unique properties of medium-resolution DEMs such as TanDEM-X and Cartosat-1 DEMs, the proposed DEM fusion techniques are used for improving the height accuracy and quality of input DEMs. Moreover, the potential of using height information produced by SAR-optical stereogrammetry is investigated as another possible source of elevations for LOD1 building modeling.

1.4.2 DEM Fusion

Data fusion approaches with many applications in remote sensing can be employed for DEM fusion tasks [17]. To this end, various methods have been investigated for different kinds of DEMs.

Among all DEM fusion methods, WA is frequently used for DEM fusion purposes because of its simple implementation and low computational cost. Benefits of the WA-based fusion of multi-sensor-derived DEMs such as an optical-derived DEM along with Interferometric SAR (InSAR) data were demonstrated in [18]. In this research, firstly, a DEM was produced from stereo SPOT images, and then the achieved DEM was employed for phase unwrapping during the InSAR DEM generation. Finally, the both DEMs were fused using weights correspondingly derived from coherence estimations for InSAR and local image correlation for optical images. Schultz et al. applied self-consistency measures for detecting outliers in optical DEMs and then fused elevations using WA [19]. Reinartz et al. [20] employed WA for the fusion of SPOT-5 and SRTM DEMs. In another study [21], WA was used to fuse ERS TanDEM data and SRTM data with MOMS-2P data. The potential of a global DEM generation by the WA fusion of SRTM data and ERS TanDEM data was investigated in [22]. Yet another study investigated the fusion of SRTM-X and -C bands using WA with weights estimated based on relative discrepancy analysis [23].

In addition to WA, most advanced techniques also apply weights to assist the fusion process in order to reach the desired output. This means that weights play a crucial role for the efficient fusion of DEMs, especially in the case of multi-sensor DEM fusion such as, optical stereoscopic-derived and InSAR DEMs [9]. In the investigations mentioned earlier, weights are mostly computed from values delivered as HEMs, which are produced by error-propagation analysis through the DEM generation process. However, evaluations demonstrate that the HEMs do not reflect all errors existing in input DEMs and cannot always lead to a successful DEM fusion [9]. Moreover, HEMs are not always available especially for optical-derived DEMs. One study used the prior knowledge of DEM qualities along with HEMs for the WA-based fusion of multi-sensor-derived DEMs such as TanDEM-X and Cartosat-1 DEMs over urban and non-urban areas [24], in which only TanDEM-X DEM was improved over non-urban areas. This signifies that even with prior knowledge, a more advanced approach should be devised to produce appropriate weights with respect to each type of input DEM particularly for multi-sensor DEM fusion using WA. In this research, an innovative and sophisticated framework is designed and implemented for producing appropriate weights usable in the WA-based fusion of multi-sensor-derived DEMs such as TanDEM-X and Cartosat-1 DEMs over urban areas.

In recent investigations, the WA-based DEM fusion has been applied for fusing multi-modal TanDEM-X raw DEMs such as ascending and descending pass DEMs [25]. Gruber et al. designed a fusion pipeline based on WA for operational mosaicking of multiple TanDEM-X acquisitions. In addition to WA, some logic for clustering consistent heights and upgrading weights regarding the influences of other significant factors such as HoA, phase unwrapping methodology and pixel locations relative to the border of the DEM scene is considered. The aim is to reach the target relative accuracy and minimize Phase Unwrapping (PU) errors remaining from initial steps [26]. While results demonstrate that the designed WA-based DEM fusion procedure can realize the target DEM, more inspections indicate that the produced TanDEM-X is not perfect in urban areas and a more advanced fusion method should be carried out for multi-modal TanDEM-X raw DEM fusion[27].

A more advanced DEM fusion technique was proposed by Papasaika [28], in which sparse representation supported by weights served for the fusion of DEMs from various data sources. The weight map for each input DEM was generated by the geomorphological properties of terrain and the nominal accuracy of the study DEM [29]. While the proposed method in [28] can finally fuse input DEMs successfully, it is still a more expensive fusion approach and strictly depends on the prior knowledge of input DEM qualities. Zach et al. implemented a TV-L_1 model for VHR multi-modal range image fusion [30]. The main drawback of TV-L_1 is a staircasing phenomenon which appears in the results of the fusion of VHR DEMs and range images. For removing this adverse effect, Pock et al. [31] proposed the Total Generalized Variation (TGV) method for the fusion of multi-modal, airborne, optical-stereoscopic DEMs, which could ultimately produce a polished DEM. A weighted version of TV and TGV was examined by Kuschk et al. [32] on different multi-modal, multi-sensor, spaceborne, optical DEMs. The weighted TGV could favorably remove noise in VHR optical-derived DEMs, but the overall performance of the fusion method decreased for multi-sensor DEM fusion. Fuss et al. utilized the modified K-means clustering algorithm to fuse multiple overlapping radargrammetric Envisat-2 DEMs [33]. While the proposed method needs lower computational load than TV-based models, its application is limited to flat areas.

Most of the mentioned advanced methods, especially TV-based models, have shown excellent performance for VHR multi-modal DEM fusion tasks. However, no study has assessed the efficiency of these types of models for fusion of the low-resolution, multi-modal DEMs such as TanDEM-X raw DEMs. On the other hand, as explained earlier, WA-based fusion does not perform flawlessly in urban areas. Thus, this research investigates the potential of TV-based models for multi-modal TanDEM-X raw DEM fusion over urban areas.

1.4.3 3D Reconstruction from SAR-Optical Imagery

In the recent decade, launching high and very high-resolution spaceborne optical and SAR sensors such as WorldView-1,2,3, TerraSAR-X, etc. has provided the possibility of 3D reconstruction of urban areas. Regarding the specific properties of each sensor type and its advantages and drawbacks, multi-sensor data fusion with an application of 3D reconstruction can be applied to benefit by integrating their instructive characteristics and reducing their defects. For example, SAR imagery with the characteristic of weather-independent imagery provides the possibility of absolute geolocalization with higher accuracy in comparison to optical imagery [34]. Moreover, a relatively perfect radiometric illumination as well as the possibility of zero nadir viewing by optical imagery makes it appropriate for stereogrammetry. Furthermore, the enormous available archive of high-resolution SAR imagery such as TerraSAR-X and growing and updating it with new data-takes in a short period as well as the archive of high-resolution optical imagery provides an opportunity to investigate pipelines for generating 3D spatial information by multi-sensor fusion.

One principal methodology for 3D reconstruction from high-resolution images is stereogrammetry, in which the 3D spatial information is produced by the intersection of rays coming from conjugate image points located in at least two stereo images with sufficient overlaps. Mostly, high-resolution optical stereo images are employed for highly accurate stereogrammetric 3D reconstruction of urban areas because of their relatively desired imaging geometries and also great radiometric representations of urban scenes. High-resolution SAR images are mostly applied for 3D reconstruction using phase information rather than gray values [35]. However, some investigations demonstrate the possibility of stereogrammetric 3D reconstruction from SAR imagery, namely radargrammetry, mostly for non-urban areas with a focus on specific applications such as phase unwrapping [36, 37]. Regarding the properties of SAR and optical imagery, grayscale images derived by the both sensor types can be employed to produce spatial information through the dense matching of SAR-optical image pairs.

The initial idea for stereogrammetric 3D reconstruction from SAR-optical imagery was provided by Bloom et al. [38]. They applied low-resolution images, SIR-B and Landsat-4/5 images for 3D reconstruction of rural areas using simplified stereophotogrammetric equations. Similarly, efforts were made for 3D reconstruction with a focus on rural areas utilizing low-resolution Seasat and SPOT/Landsat images [39, 40, 41]. These studies presented more complicated rigorous sensor models for the applied SAR and optical imagery and finally used a block adjustment pipeline to produce a coarse DEM in non-urban areas. In another study, Xing et al. used ERS-2/Radarsat-1 and SPOT data to produce 3D information [42]. They designed a block adjustment tool employing the co-linearity equations for spaceborne optical imagery and the range-Doppler equations for SAR imagery. They could finally estimate the positions of ground points by solving the defined adjustment equations. While the accuracy of all investigations as mentioned earlier lies in the domain of dekameter, Wegner et al. estimated building footprints with an accuracy of several meters from a single-pass InSAR image pair and aerial imagery but not using SAR-optical stereogrammetry [43]. The challenges and possibility of SAR-optical stereogrammetry from high-resolution imagery were investigated in [44, 45]. An object-based strategy applying the SIFT similarity measure in a template-based matching was used in [44] to explore the heights of target points. More similarity measures were assessed in [45]. This study also proposed a strategy for simultaneous tie point matching and 3D reconstruction, which exploits an epipolar-like search window constraint. Moreover, the study discussed the effects of SAR-optical intersection geometry and acquisition configurations on the final accuracy of 3D reconstruction.

As illustrated earlier that relatively few studies have dealt with 3D reconstruction from SAR-optical image pairs, there has yet been no investigation into the feasibility of a dense multi-sensor stereo pipeline as known from photogrammetric computer vision. Thus, this research investigates the possibility and potential of urban 3D reconstruction and height generation from VHR SAR-optical imagery such as TerraSAR-X and WorldView-2 by implementing a dense matching pipeline.

2 Fundamentals

This chapter provides some necessary information and fundamentals used in this Ph.D. research, which include the LOD1 building modeling process, the initial concepts of multi-sensor data fusion, an introduction to DEMs and specifically short descriptions of TanDEM-X and Cartosat-1 DEMs, prerequisites and essential preprocessing for DEM uncertainty assessment and also basics of DEM fusion. Furthermore, some materials and principles are presented, such as those applied for SAR-optical stereogrammetry. Useful references are also introduced to provide opportunities for enthusiasts to go into the depth of the presented information and concepts.

2.1 LOD1 3D Reconstruction

The main objective of this dissertation is to investigate the possibility of LOD1-based 3D building modeling from different remote sensing data sources which can be efficiently applied to wide areas. For this purpose, the heights improved by multi-sensor and multi-modal DEM fusion techniques or produced by SAR-optical stereogrammetry will be used for 3D building modeling and finally prismatic model generation. Due to employing the medium resolution of the input DEMs and consequently insufficient details for building outline detection, only LOD1 models can be reconstructed from those heights. Also, as will be shown in Section 3.2.2, the point cloud resulted from SAR-optical stereogrammetry is partially sparse and as a result, building outlines cannot be recognized.

One popular option is to exploit the building footprint layer provided by OSM. Then, the heights of building outlines can be derived from either those fused DEMs or the point cloud achieved by SAR-optical stereogrammetry. Technically, this can be realized in two steps. The first step is to classify heights to those located inside and outside building outlines. Then, only points that are within building outlines are kept while the remaining points are discarded. After that, for each remaining height, the Id of the corresponding building (in which the height is located) is assigned. It facilitates the process of joining building footprints layer to heights.

There are several elevation references that should be considered for estimating an ultimate height for a building outline [46]. These references are displayed in Figure 2.1. 3D reconstruction based on those levels can be realized by using high-resolution data such as LiDAR point clouds along with precise cadastral maps. It should be noted that among different vertical references, the optimal LOD1 building modeling can be realized by choosing the half of the roof height as elevation reference [47]. Specifying this level in medium resolution remote-sensing-derived heights, however, is not possible. Therefore, for LOD1 3D building reconstruction using medium resolution data such as those applied in this research, only median or mean of heights inside a building outline is used. The main advantage of the median is its

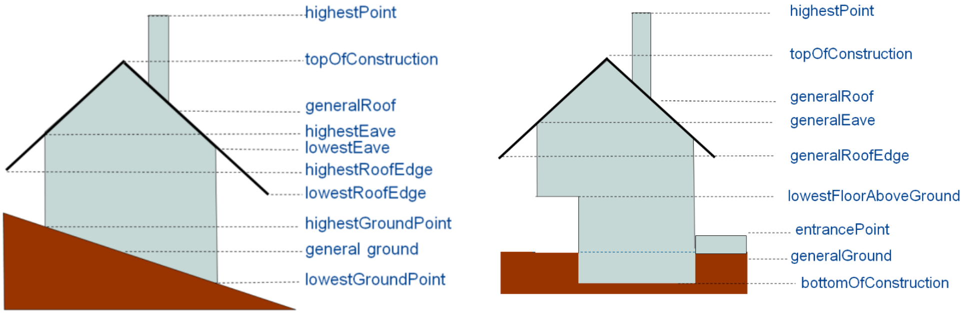


Figure 2.1: Examples of elevation references for different kinds of buildings [46]

robustness against outliers in comparison to the mean measure. Finally, LOD1 model can be produced by reconstructing each building as a simple box using its outline and the median-based allocated height.

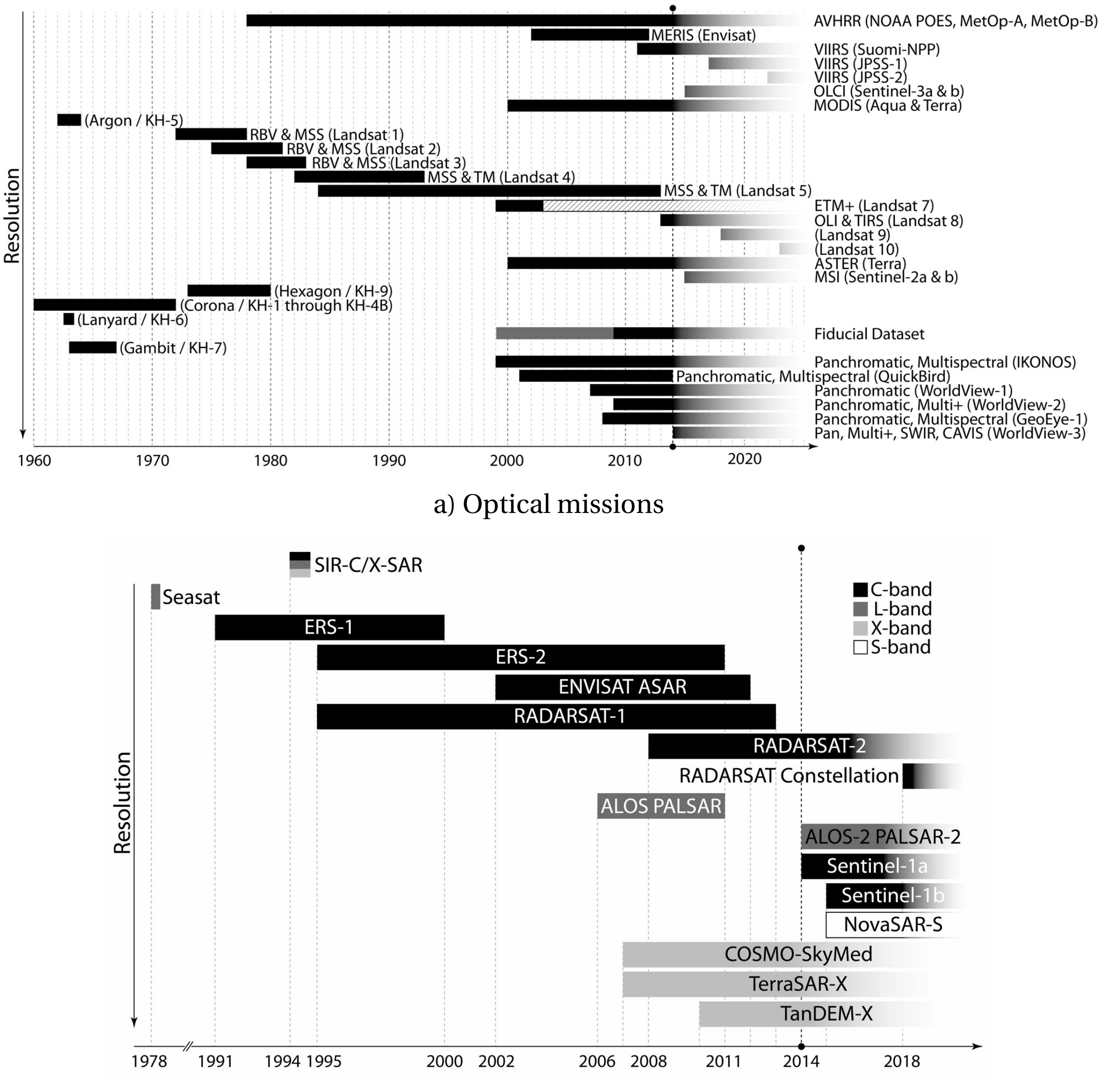
2.2 Data Fusion in Remote Sensing

Today, a growing number of satellites equipped with various kinds of sensors provide a huge archive of various remotely-sensed data and images of our planet with different specifications regarding resolution, accuracy, coverage and spectral imaging ability, etc. Among different types of remote sensing sensors, SAR and optical data are more popular because of their broad ranges of applications in different fields. For example, the TanDEM-X mission provides high-resolution bistatic SAR images for the whole earth. Moreover, a large coverage of landmasses is enabled by the Sentinel-1, 2 and 3 missions. 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, extensive archives of satellite imagery acquired by different sensors are available and will not stop growing in the future. Figure 2.2 confirms an increasing number of different types of remote sensing sensors in space.

Each kind of sensor has its distinct properties, e.g. wavelength, resolution, accuracy, and coverage, etc. For instance, the Sentinel-1 mission provides a global, cloud-free medium-resolution SAR dataset, while the Sentinel-2 mission provides easy-to-interpret multi-spectral data that is well-suited for land-use/land-cover mapping tasks, the acquired data is profoundly affected by cloud coverage.

Data fusion can be applied for integrating datasets with different specifications to enhance information extraction by beneficially combining the properties of individual sensors [17, 49].

In remote sensing applications, data fusion mainly includes several steps. The first step is data alignment which mainly refers to the spatial and temporal coregistration of sensor data. After that, data is correlated and commonly achieved by resampling. To this end, a reference frame is defined and all input data are transformed into the defined reference frame by an interpolation algorithm. Then, appropriate identities or attributes depending on data and the fusion objective such as 3D coordinates, size or area are extracted as features. At the final



a) Optical missions

b) SAR missions

Figure 2.2: Increasing number of remote sensing satellites in space [48]

step, extracted features are integrated depending on types of sensors [17].

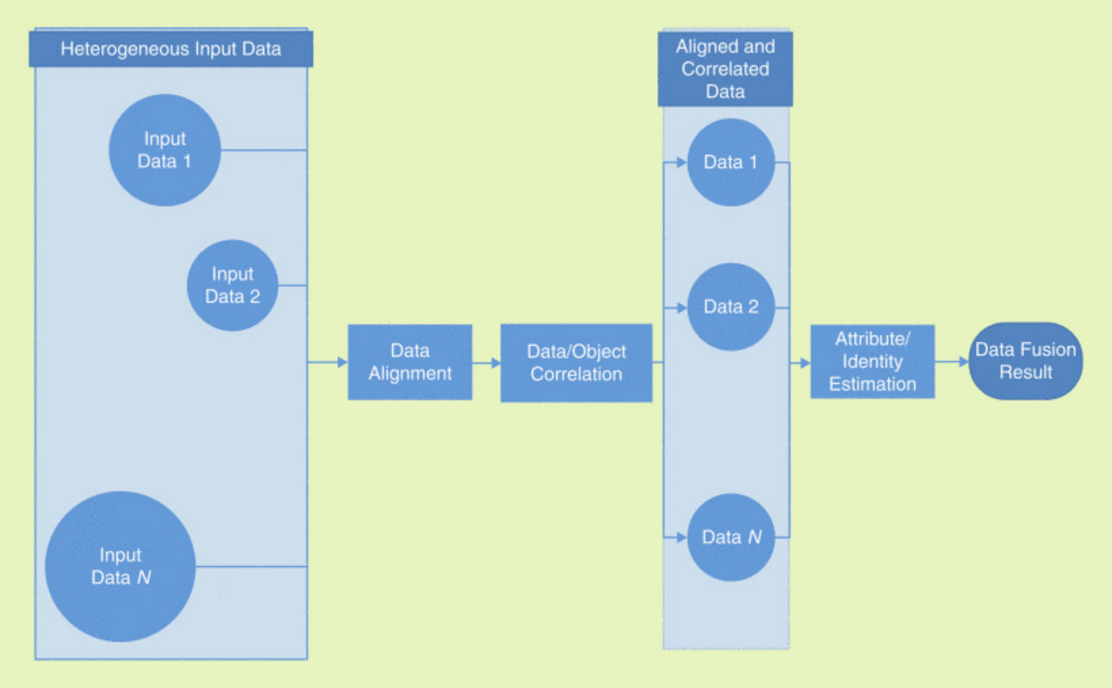


Figure 2.3: Flowchart of the generalized data fusion system [17]

A flowchart of the generalized fusion system consisting of the described fusion steps is depicted in Figure 2.3. For example, in this research, "Data Alignment" is performed for aligning input DEMs in DEM fusion as well as for coregistration of optical and SAR imagery in SAR-optical stereogrammetry. Moreover, "Data Correlation" is performed by image matching in SAR-optical stereogrammetry and by fusion techniques in DEM fusion to finally estimate heights in the "Attribute Estimation" module.

2.3 Digital Elevation Models

DEMs in different resolutions, levels of height accuracy and coverages are routinely produced by different techniques for a varied range of applications in different fields, such as navigation, geographical studies of the environment, or the ortho-rectification of remote sensing imagery.

Particular attention is paid to the production of global DEMs, which represent homogeneous topography information for nearly all landmasses of the world. Different technologies have been employed for producing nearly global DEMs like the SRTM DEM [50],[51], the ASTER GDEM [52] or AW3D30 [53, 54] which lie in two categories: SAR-interferometric and optical stereoscopic procedures. Each one of them has its own advantages and drawbacks that lead to DEMs with specific properties and limitations regarding final resolution and coverage. As an example, the SRTM DEM with a grid spacing of $1''$ only covers the latitudes between 56°S and 60°N. An example for an elevation model derived from optical stereo data is the AW3D30 DEM based on ALOS PRISM data, which provides both higher accuracy and larger coverage (between 82°S and 83°N) than the SRTM DEM, but contains some void areas due to missing information caused by clouds, snow etc. [55].

Recently, a new global topography dataset was attained through the TanDEM-X mission, which provides a spatial resolution of 12m with coverage of nearly the whole earth. In the following, the main characteristics of the TanDEM-X DEM are presented. After that, the properties of Cartosat-1 DEM which will be used for the quality improvement of TanDEM-X DEM are also introduced.

2.3.1 TanDEM-X DEM

The TanDEM-X mission realized a new global DEM covering almost the whole planet. The TanDEM-X mission comprises twin SAR satellites (TerraSAR-X and TanDEM-X launched in June 2007 and June 2010, respectively), which fly in adjacent orbits to acquire bistatic SAR images. The mission was devised to produce DEMs with a target accuracy according to High-Resolution Terrain Information standard level 3 (HRTI-3) [56]: i.e., with a relative height accuracy better than 2 m for areas including slopes lower than 20%, and 4 m for slopes steeper than 20% [10].

The particular satellite constellation equipped with X-band SAR sensors exploits a bistatic SAR interferometry configuration with single pass acquisitions free of atmospheric and temporal decorrelation effects and consequently provides the first high-resolution global DEM.

Form raw SAR data takes to the final global DEM, a workflow including different phases such as interferogram generation, phase unwrapping, data calibration, DEM block adjustment, and mosaicking is implemented at German Aerospace Center (DLR) [57]. A primary step of the DEM generation procedure is carried out in the Integrated TanDEM-X Processor (ITP). The initial DEM product, the so-called raw TanDEM-X DEM with nominal pixel spacing of 0.2 arcsec (6 m at the equator), is the output of ITP [58, 59].

During the raw DEM generation, some potential error sources are removed by instrument and baseline calibration [60]. After that, the vertical bias which usually lies between 1 m to 5 m is corrected by the least squares block adjustment [61]. The block adjustment is performed by using ICESat data and connecting points in the overlapping areas of raw DEM tiles. However, dependent on the terrain morphology, some error sources remain after the block adjustment. The effect of these errors can be decreased through the fusion of several DEM coverages within DMP [62].

The TanDEM-X raw DEM coverage over different terrain types is displayed in Figure 2.4. As can be seen, most of the world is covered by at least two nominal acquisitions with Height of Ambiguities (HoA) between 30 m and 55 m. The main objective of TanDEM-X DEM fusion is to improve the final accuracy by employing several coverages over different areas [26].

The raw TanDEM-X DEM is finally cast in a grid with a pixel spacing of 0.4 arcsec (12 m at the equator) after DEM fusion to obtain the global DEM according to HRTI-3 standard [26].

While the standard DEM globally represents non-urban areas with unprecedented relative accuracy [63], the drop of the DEM's spatial resolution makes the final standard DEM unsuitable for high-resolution 3D reconstruction in urban areas [64]. Consequently, the raw

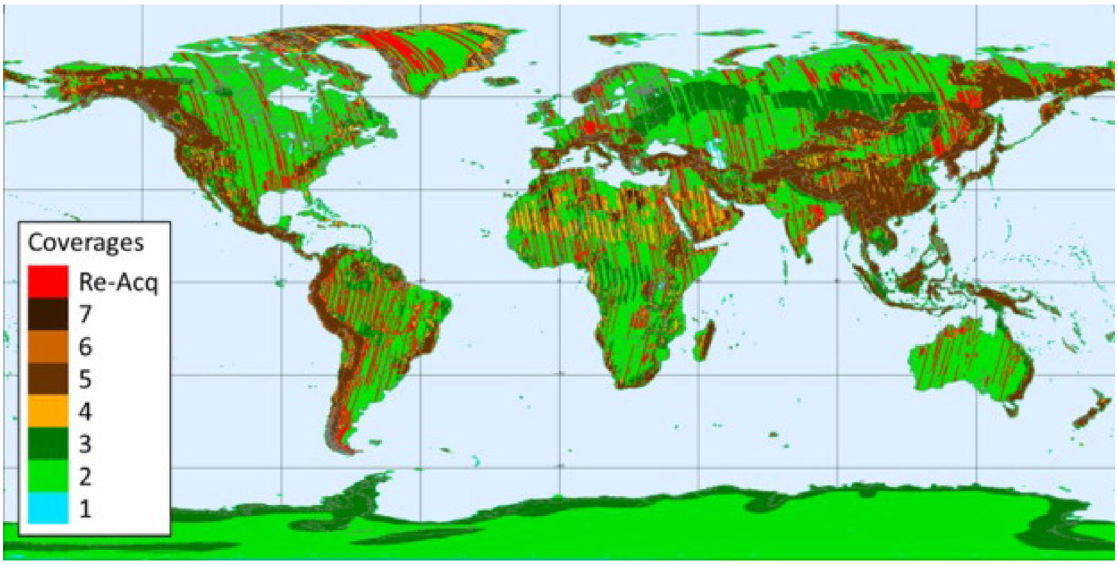


Figure 2.4: TanDEM-X coverage in different areas [57]

TanDEM-X DEM provides a more spatially detailed mapping of urban areas in comparison to the standard version of the global TanDEM-X DEM.

2.3.2 Cartosat-1 DEM

Cartosat-1 (also called IRS-P5) is an Indian satellite (launched in May 2005) equipped with a pushbroom sensor consisting of an ensemble of CCDs with a size of 2.5 m in two lines for along track scanning of scenes with a stereo angle of 3 1° [65].

Cartosat-1 data provides a series of DEMs with the relative accuracy of HRTI-3 standard (2 −3 m). It is particularly intended to produce a high-resolution DEM with coverage of a relatively wide area [66] and is used, for instance, for large-scale DEM generation in Europe [67]. Figure 2.5 illustrates the capability of using the Cartosat-1 stereo images for a large scale DEM generation [68].

The Cartosat-1 data are provided with Rational Polynomial Coefficients (RPCs) computed from the mission's orbit and attitude information. Evaluations have demonstrated that their accuracy is restricted to multiple hundred meters [69], i.e. the final produced DEM despite fairly high relative accuracy is absolutely located in an incorrect position. The poor accuracy of the RPCs affects the stereo intersection results and causes residuals in the final DEM product.

Generally, a proper distribution of Ground Control Points (GCPs) is needed for RPC refinement and bias compensation [70] of high-resolution optical images like those provided by Cartosat-1, but the availability of GCPs cannot always be ensured. The conventional solution is to use available global DEMs—like the SRTM DEM —as an external vertical reference for bias compensation and RPC refinement [71]. The process of RPC correction is depicted as a diagram displayed in Figure 2.6.

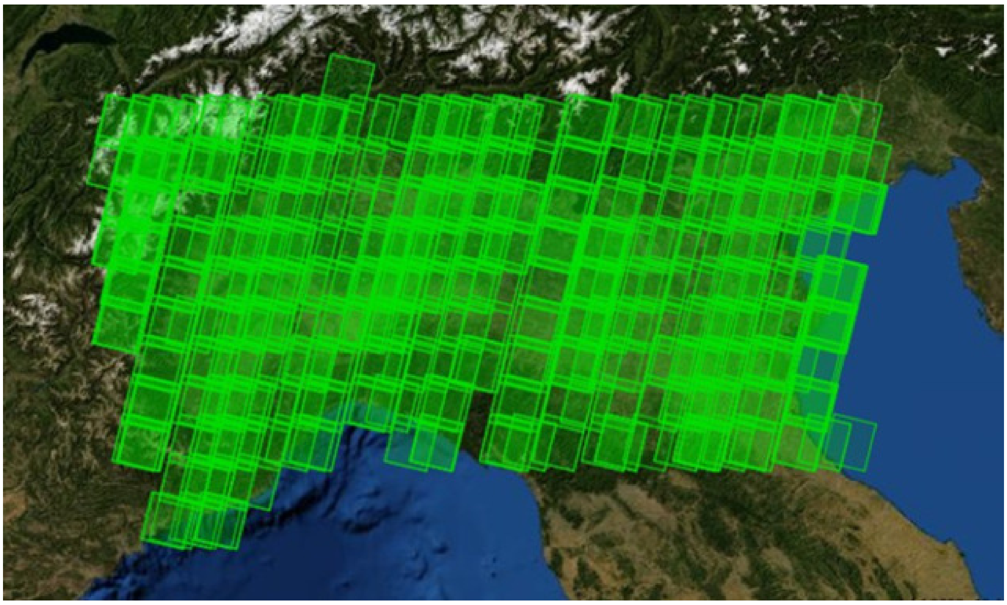


Figure 2.5: Stack of stereo Cartosat-1 images acquired over the north of Italy for producing a DEM on a large scale [68]

As illustrated in Figure 2.6, RPCs of Cartosat-1 imagery can be corrected in two steps. First, the Cartosat-1 imagery is aligned to reference imagery to generate required tie points and GCPs. Then, the horizontal bias compensation performed by a preliminary RPC correction using achieved GCPs. For instance, Landsat ETM+ or Sentinel-2 imagery can be applied as reference horizontal data for large scale mapping purposes. For DEM generation over a local area especially urban areas, the highly accurate horizontal data such as aerial orthophotos are suitable choices. In the next step, the vertical bias is compensated using an external DEM such as SRTM [68].

After RPC correction and bias compensation during a block adjustment, the final DEM is produced by implementing a dense matching algorithm. More details of Cartosat-1 DEM generation can be found in [72].

2.3.3 DEM Uncertainty Assessment

For uncertainty assessment, the differences of vertical and horizontal datums between the test DEMs and the reference DEM should be removed. Therefore, all data must first be transferred to one reference datum with identical pixel spacing.

After datum homogenization, the test DEMs must be precisely aligned to the reference DEM in order to remove any rotational and horizontal translations. Several algorithms such as Iterative Closest Point (ICP) algorithm [73], least square matching [74], or manual registration can be used for DEM coregistration.

After data preparation, the DEMs can be evaluated with respect to the reference dataset. The accuracy of the DEMs is evaluated in terms of absolute and relative accuracy with re-

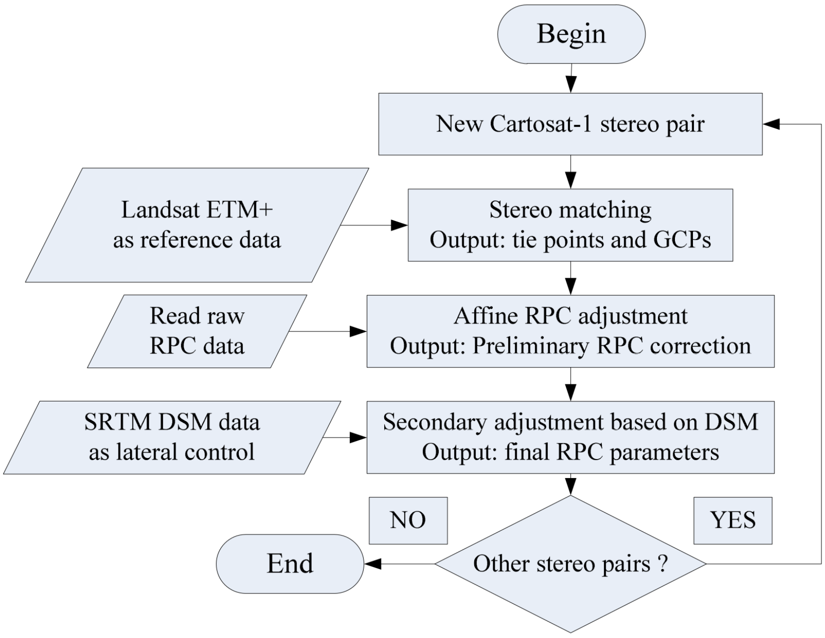

Begin
New Cartosat-1 stereo pair
Landsat ETM+ as reference data
Stereo matching Output: tie points and GCPs
Read raw RPC data
Affine RPC adjustment Output: Preliminary RPC correction
SRTM DSM data as lateral control
Secondary adjustment based on DSM Output: final RPC parameters
NO
YES
End
Other stereo pairs ?

Figure 2.6: Two-step process of the Cartosat-1 RPC correction [68]

spect to the LiDAR data. The absolute vertical accuracy is indicated by the average vertical differences between the compared DEMs and the ground truth while the relative accuracy refers to pixel-wise height precision and undulation.

The usual metrics for precision evaluation are Root Mean Square Error (RMSE) and LE90 (Linear Error in 90% confidence interval; a common metric in the TanDEM-X specification document to express TanDEM-X DEM accuracy locally and globally [75]). LE90 is calculated based on the Standard Deviation (LE90=1.645 × STD) in case that height residuals follow a normal distribution. However, in general, one can not assure the underlying normal distribution for errors. In addition, if the quantile-quantile plot of the height errors over any study area demonstrates that the errors do not follow a normal distribution, other robust metrics like Normal Median Absolute Deviation (NMAD) are recommended for error analysis [76] and LE90 is computed by 1.645 × NMAD. Finally, the RMSE is gained after outlier removal by marginal LE90 as a threshold.

2.4 DEM Fusion Basics

Similar to the uncertainty assessment of DEMs, the datum homogenization, coregistration and subsequently data correlation should be performed before implementing any DEM fusion pipelines. For stability reasons, in addition, the height data should be normalized to the interval [0, 1] [32]:

$$h_k^n(x,y) = \frac{h_k(x,y) - h_{min}}{h_{max} - h_{min}}, \tag{2.1}$$

where $h_k(x,y) > 0$ is the elevation of the study DEM with index k at location (x,y), $h_{max} > 0$ and $h_{min} > 0$ $(h_{min} < h_{max})$ are the lowest and highest elevations among all input DEMs. The output gives the normalized height in the considered location.

The most famous, high-speed and low computational cost method for DEM fusion is WA which is implemented by

$$\mathbf{f} = \sum_{i=1}^{k} \mathbf{w}_i \odot \mathbf{h}_i, \tag{2.2}$$

where $\mathbf{h}_i$ are 2D arrays representing the input DEMs, $\mathbf{w}_i$ are the corresponding weight maps and $\odot$ is a pixel-wise product. It is worth to note that other simple methods such as pixel-wise median or mode based fusion can also be employed for DEM fusion especially when multiple DEMs are available [77].

The main critical issue for using WA for DEM fusion is to apply appropriate weights that are fairly representative of expected height errors in the source DEMs. For instance, for TanDEM-X DEM mosaicking and multiple raw DEM fusion, generally, these weights are delivered as HEMs from the ITP. For each height of the TanDEM-X DEM, the corresponding HEM value can be estimated by

$$\sigma_j = H_{amb} \frac{\sigma_{\phi,j}}{2\pi}, \tag{2.3}$$

where H_{amb} is the height of ambiguity and $\sigma_{\phi,j}$ is the interferometric phase error that is estimated from the interferometric coherence and the InSAR geometry [10]. Then, from these values, the respective weights can be calculated for each pixel location by

$$w_j = \frac{\frac{1}{\sigma_j^2}}{\sum_{j=1}^{N} \frac{1}{\sigma_j^2}}. \tag{2.4}$$

For optical-derived DEMs such as Cartosat-1 DEM, the standard deviation of the stereo matching process can be used to produce a similar HEM.

2.5 Fundamentals of Stereogrammetry

Conventionally, 3D reconstruction in remote sensing is either based on exploiting phase information provided by InSAR, or on space intersection in the frame of photogrammetry with optical images or radargrammetry with SAR image pairs. In all these stereogrammetric approaches, at least two overlapping images are required to extract 3D spatial information. For this aim, a 3D reconstruction framework is implemented. The main stage of this framework is to carry out a dense matching algorithm for 3D reconstruction, but before that, some primary processes such as establishing epipolarity constraint and implementing a block adjustment are fulfilled. Finally, the 3D reconstruction is realized by a forward intersection and using disparities achieved through a dense matching process.

2.5.1 Epipolarity Constraint

In most stereogrammetric 3D reconstruction scenarios, the epipolarity constraint facilitates the procedure of image matching by reducing the search space from 2D to 1D [78]. The epipolarity constraint always exists for optical stereo images captured by frame-type cameras that follow a perspective projection [79]. This phenomenon is illustrated in Fig. 2.7.

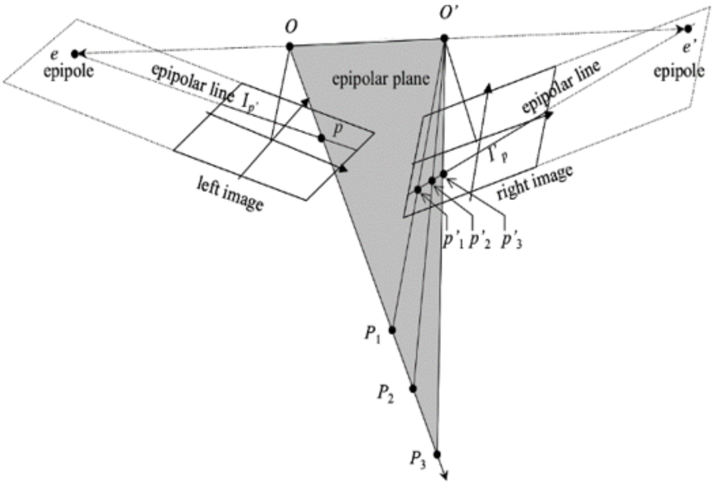


Figure 2.7: Epipolarity constraint for frame-type camera [78]

For a point, p in the *left-hand* image, the conjugate point in the corresponding *right-hand* image is located on the so-called epipolar line. This epipolar line lies on the plane passing through both image projection centers (O, O') and the image point p. It can also be obtained by changing the depth or height of p in the reference coordinate system. While it is known that epipolar lines exist for images captured from frame-type cameras, straightness cannot be ensured for other sensor types [79]. Thus, epipolar curves are referred instead of epipolar lines to express generality in the remainder of this dissertation. In general, epipolar curves in image pairs captured by frame-type cameras (as shown in Fig. 2.7) can be described as [80]

$$l_r = \mathbf{F}^T p', \tag{2.5}$$

where l_r refers to the epipolar curve in the right-hand image associated with the image point p' on the left-hand image. $\mathbf{F}$ is the fundamental matrix, which includes interior and exterior orientation parameters for projecting coordinates between the two images. Similarly, an epipolar curve in the left-hand image can be written as $l_l = \mathbf{F}p''$.

Since in remote sensing, optical satellites are usually equipped with linear array pushbroom sensors instead of frame type cameras, the epipolarity constraint becomes more complicated. Consequently, the fundamental matrix for push-broom sensors is more complicated than that for frame-type sensors. This is caused by the fact that there is no unique projection center for the whole acquired scene. In addition, the satellite trajectory tracking is not as simple as for frame-type airborne platforms [81, 82, 83]. For push-broom satellite image pairs, the epipolarity constraint can be verified similarly, but linear arrays are substituted for a frame image.

With respect to remote sensing, several studies have demonstrated that the epipolar curves for scenes acquired by linear array push-broom sensors are not straight [81, 84]. For example, Kim [84] used the model developed by Orun and Natarajan [85] to prove that the epipolar curves in SPOT scenes look like hyperbolas. Orun and Natarajan's model assumes that the rotational roll and pitch parameters are constant during the flight, while time-dependent quadratic polynomials can model the yaw. Morgan et al. [86] demonstrated that the epipolar curves would not be straight even with uniform motion.

For a SAR sensor, the imaging geometry is entirely different from that of optical sensors, as data are collected in a side-looking manner based on the range-Doppler geometry [87]. However, the possibility of establishing the epipolarity constraint in stereo SAR image pairs has been investigated by Gutjahr et al. [37] and Li and Zhang [88] for radargrammetric 3D reconstruction. Gutjahr et al. experimentally showed that epipolar curves in SAR image pairs are also not perfectly straight, but can be approximately assumed to be straight for radargrammetric 3D reconstruction tasks through dense matching [37].

2.5.2 Block Adjustment

Before carrying out the dense matching process, a block adjustment approach should be implemented to align the corresponding images to a reference image. The output of the block adjustment will lead to relatively bias compensation of the corresponding images. This process can be performed using a block adjustment which is based on RPCs instead of rigorous sensor models as proposed in [89] and consequently RPCs of corresponding images are modified. Generally, designing an appropriate function for modeling the existing bias in the RPCs given by the optical image depends on the sensor properties [90], but for most sensors, an affine model can be applied [91]. The affine model for RPC bias compensation can be formulated as

$$\begin{aligned}\Delta x &= m_0 + m_1\, x_o + m_2\, y_o \\ \Delta y &= n_0 + n_1\, x_o + n_2\, y_o\,,\end{aligned} \tag{2.6}$$

where x_o,y_o represent column and row of tie points in the corresponding images and m_i and n_i $(i = 0, 1, 2)$ are unknown affine parameters to be estimated through the block adjustment procedure.

For implementing block adjustment, tie points are selected between a reference image and a corresponding image. The tie points can either be selected by a sparse key point matching method or manually. In the end, the block adjustment equations can be constituted and solved by least squares.

2.5.3 Dense Matching: SGM

The core step in a stereogrammetric 3D reconstruction workflow is the dense image matching algorithm to obtain the disparity map, which can then be transformed into the desired 3D point cloud. Generally, two different dense matching rationales can be used according to whether local or global optimization is more important [92]. For the case of global optimization, an energy functional consisting of two terms is established to find the optimal disparity map [93]:

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d), \tag{2.7}$$

where $E_{data}(d)$ is a fidelity term that makes the computed disparity map consistent with the input image pairs, $E_{smooth}(d)$ considers the smoothness condition for the disparity map, and λ is a regularization parameter that balances the fidelity and smoothness terms.

For a given image pair, the disparity map is calculated by minimizing the energy functional in (2.7). The main advantage of global dense matching over local matching methods is greater robustness against noise [92], although most existing algorithms for global dense image matching have a higher computational cost [94].

SGM as a well-known dense matching method devised by [94], offers acceptable computational cost and high efficiency and performs very similarly to global dense image matching.

In SGM, the disparity map is computed by minimizing a cost function, defined as a data term, over the whole image along paths toward the target pixel. The global energy functional for SGM can be expressed as

$$S(\mathbf{p}, d) = \sum_{\mathbf{r}} L_{\mathbf{r}}(\mathbf{p}, d), \tag{2.8}$$

where $L_{\mathbf{r}}$ includes a cost function and two penalties, one smaller value P_1 for pixels where the disparity difference is small and another higher value P_2 for pixels where the disparity difference is larger. The costs along different predefined paths toward the target pixel $\mathbf{p}$ are then aggregated to generate a semi-global energy. L_r is configured as

$$L_{\mathbf{r}}(\mathbf{p}, d) = C(\mathbf{p}, d) + min \left\{ \begin{array}{c} L_{\mathbf{r}}(\mathbf{p}-\mathbf{r}, d), \\ L_{\mathbf{r}}(\mathbf{p}-\mathbf{r}, d-1) + P_1, \\ L_{\mathbf{r}}(\mathbf{p}-\mathbf{r}, d+1) + P_1, \\ \min_i L_{\mathbf{r}}(\mathbf{p}-\mathbf{r}, i) + P_2 \end{array} \right\}. \tag{2.9}$$

Thus, for target pixel $\mathbf{p}$ at disparity d and in direction r, the cost function is calculated using a similarity measure. The equation also adds the minimum cost of the previous pixel $\mathbf{p}-\mathbf{r}$ including the penalties P_1, P_2 for the nearest and farthest disparities to smooth any discontinuities in the final disparity map.

To determine the optimal disparity map, a hierarchical matching procedure is adopted to rapidly estimate the higher-level disparity maps from the initial disparity values. In each level (from top to bottom), the disparity map computed by minimizing $\{S(\mathbf{p}, d)\}$ in the previous level is employed as the initial disparity map in the current level. At the lowest level, the final disparity is estimated for the left-hand image.

2.5.4 Similarity Measure

As described in the previous section, the SGM cost function measures the cost of candidate conjugate points to achieve the optimum disparity map. In the following, a great cost function, applied in this work, namely Mutual Information (MI) is described. MI is formed based on the entropies of the source images by

$$MI(I,I') = H(I) + H(I') - H(I,I'), \tag{2.10}$$

where $H(I)$ and $H(I')$ are the entropies of the source images I and I', and $H(I,I')$ is the joint entropy of the two images. In a discrete space like image, theses entropies and the joint entropy can be defined as

$$H(I) = \sum_{i} -\frac{1}{n}(\log p_I(i) * g(i)) * g(i) \tag{2.11a}$$

$$H(I') = \sum_{i'} -\frac{1}{n}(\log p_{I'}(i') * g(i')) * g(i') \tag{2.11b}$$

$$H(I,I') = \sum_{i,i'} -\frac{1}{n}(\log p_{I,I'}(i,i') * g(i,i')) * g(i,i'), \tag{2.11c}$$

where $p_I(i)$ and $p_{I'}(i')$ are the marginal probability density functions and $p_{I,I'}(i,i')$ is the joint probability density function. $*g()$ is the Gaussian convolution that is applied to the entropy and joint entropy functions. Matched points are points with higher MI information that can become through minimizing the joint entropy.

2.5.5 Sensor Geometry

In the process of 3D reconstruction, the sensor geometry of each applied imagery has a significant role in different steps such as establishing epipolarity constraint, block adjustment and finally forward intersection for generating 3D spatial information from disparity maps. Since this dissertation investigates the possibility of 3D reconstruction from SAR-optical image pairs, the geometries of spaceborne SAR and optical sensors are introduced in the following.

Most of the spaceborne optical imagery especially VHR images are collected by linear array push-broom sensors. A collinearity condition can be used to formulate a rigorous model for reconstructing the imaging geometry of a linear array push-broom sensor. This rigorous model can be expressed by [95]

$$\begin{pmatrix} x_l = 0 \\ y_l \\ f \end{pmatrix} = \lambda R_{\omega(t)\phi(t)\kappa(t)} \begin{pmatrix} X - X^o(t) \\ Y - Y^o(t) \\ Z - Z^o(t) \end{pmatrix}, \tag{2.12}$$

where (x_l, y_l) are the coordinates of point p in the linear array coordinate system, f is the focal length, $(X^o(t), Y^o(t), Z^o(t))$ represents the satellite position at time t in the reference coordinate system, (X, Y, Z) are the ground coordinates of the target point T, λ is the scale factor, and $\mathbf{R}_{\omega(t)\phi(t)\kappa(t)}$ is the 3D rotational matrix computed from rotations $\omega(t), \phi(t), \kappa(t)$ along the three dimensions at time t. Note that those as mentioned earlier rotational and translational components are estimated by time-dependent polynomials.

Similarly, a rigorous model describing the range-Doppler geometry [96] (displayed in Fig. 3.14 as well) can also be applied to SAR imagery. In this model, the slant-range equation is first used to describe the range sphere as [87]

$$R = \|\mathbf{R}_{CT} - \mathbf{R}_{CS}\|, \tag{2.13}$$

where R is the slant-range and $\mathbf{R}_{CT}$, $\mathbf{R}_{CS}$ are the target point and SAR sensor position vectors in the reference coordinate system. C refers to the center of the reference coordinate system. For a given pixel y_r in the slant-range SAR scene, equation (2.13) can be reformulated as

$$R = c\,t = c\,(t_0 + \frac{y_r}{2f_r}) = c\,t_0 + c\,\frac{y_r}{2f_r} = R_0 + \Gamma\,y_r, \tag{2.14}$$

where R is the slant-range of the target point, c is the velocity of light, t_0, t are the one-way signal transmission times for the first range pixel and range pixel coordinate y_r, respectively, and f_r is the range sampling rate. R_0 gives the slant-range for the first range pixel and $\Gamma = \dfrac{c}{2f_r}$.

The second equation describes the geometry of the Doppler cone:

$$f_D = \frac{2}{\lambda_r R}\mathbf{V} \cdot (\mathbf{R}_{CT} - \mathbf{R}_{CS}), \tag{2.15}$$

where f_D is the Doppler frequency, λ_r is the SAR signal wavelength, $\mathbf{V}$ is the velocity vector and $\cdot$ denotes the inner product operator.

As another possibility for describing a sensor geometry, RPCs are a well-established substitute for the rigorously derived optical imaging model. They are widely used for different purposes such as epipolar curve reconstruction [89, 97, 98, 99, 100, 101, 102, 103]. The relation between the image space and the geographic reference system is created by the rational functions [104]:

$$c = \frac{P_1(\lambda, \phi, h)}{P_2(\lambda, \phi, h)} = f(\lambda, \phi, h) \tag{2.16}$$

and

$$r = \frac{P_3(\lambda, \phi, h)}{P_4(\lambda, \phi, h)} = g(\lambda, \phi, h), \tag{2.17}$$

where r, c are normalized image coordinates, i.e. normalized rows and columns of points in the scene and ϕ, λ, and h denote the normalized latitude, longitude, and height of the respective ground point. The relationship between normalized and un-normalized coordinates is given by [105]

$$X = \frac{X_u - X_o}{S_x}, \tag{2.18}$$

where X is the normalized coordinate, X_u is the un-normalized value of the coordinate, and X_o, S_x are the offset and scale factor, respectively.

In equations (2.16) and (2.17), P_i $(i = 1, ..., 4)$ are n-order polynomial functions that are used to model the relationship between the image space and the reference system. They can be written as

$$\begin{aligned} P_i = a_{i,0} + a_{i,1}h + a_{i,2}\phi + a_{i,3}\lambda \\ + a_{i,4}h\phi + a_{i,5}h\lambda + a_{i,6}\phi\lambda + a_{i,7}h^2 + a_{i,8}\phi^2 + a_{i,9}\lambda^2 \\ + a_{i,10}h\phi\lambda + a_{i,11}h^2\phi + a_{i,12}h^2\lambda + a_{i,13}\phi^2 h + a_{i,14}\phi^2\lambda \\ + a_{i,15}h\lambda^2 + a_{i,16}\phi\lambda^2 + a_{i,17}h^3 + a_{i,18}\phi^3 + a_{i,19}\lambda^3, \end{aligned} \tag{2.19}$$

where $a_{i,n}$ $(n = 0, 1, ..., 19)$ are the polynomial coefficients.

For projection from the image space to terrain, the inverse form of the rational function models is used:

$$\lambda = \frac{P_5(c, r, h)}{P_6(c, r, h)} = f'(c, r, h) \tag{2.20}$$

and

$$\phi = \frac{P_7(c, r, h)}{P_8(c, r, h)} = g'(c, r, h). \tag{2.21}$$

For this task, another set of RPCs for inverse projection as well as the terrain height h is needed.

The RPCs for optical sensors are usually delivered by vendors alongside the image files. For SAR sensors RPCs can be estimated using ephemerids and orbital parameters attached to the data based on the terrain-independent approach [106].

3 Summary of Investigations

Regarding the objectives of this dissertation mentioned in Section 1.2, this chapter provides summaries of contributions and investigations published as peer-reviewed papers. More details can be found by respecting to the papers available in the appendices. This chapter is organized into three sections. The first section summarizes the research implemented for improving DEM quality, especially over urban areas using innovative multi-sensor and multi-modal DEM fusion techniques. Section 3.2 provides a summary of the devised framework for stereogrammetric 3D reconstruction from SAR-optical image pairs. The heights derived from DEM fusion and SAR-optical stereogrammetry are applied in the final section. This section investigates the potential of 3D building model generation using building footprints provided by OSM as VGI and those elevations produced by the methods described in the earlier sections.

3.1 Height Retrieval by DEM Fusion

A great remote sensing source that provides height information for 3D reconstruction is DEM. As explained in Section 2.3, DEMs are mainly produced from optical stereoscopy or SAR interferometry or are derived from LiDAR data. Medium-resolution DEMs such as TanDEM-X DEM are more popular as they provide the global coverage of the planet with unprecedented relative accuracy, which can potentially be applied for large-scale 3D reconstruction. However, the primary produced DEM quality is not perfect in the first place. For instance, the preliminary visual inspection of raw TanDEM-X DEM data still indicates unfavorable spatial resolution and drop of height precision, especially for areas with topographically difficult surfaces —such as urban areas [107]—and also reveals the requirement for primary produced DEM quality enhancement in these areas.

One solution for refining a DEM such as TanDEM-X DEM in difficult terrains can be a fusion with elevation data derived from other sources with different acquisition properties which do not suffer from sensor-inherent imaging effects such as layover and shadowing. Examples for these alternative data are DEMs derived from optical stereo imagery.

Another possibility to gather reliable height information is to fuse multi-modal DEM products. For instance, WA is currently applied as a fast and straightforward method for TanDEM-X raw DEM fusion, in which weights are computed from HEMs delivered from ITP. While the WA approach can realize the predefined goals in DMP for global DEM generation, it does not perform well in difficult terrains with complex morphology such as urban areas which contains many high-frequency contents such as edges. Accordingly, advanced fusion approaches can be substituted for WA in the process of DEM mosaicking.

This study focuses on the two aforementioned possible solutions: multi-sensor-based, and multi-modal-based DEM fusion strategies for improving the quality of medium-resolution DEMs —e.g. TanDEM-X DEM, —and subsequently generating highly accurate elevations for urban 3D reconstruction. Thus, this section is a summary of the two contributions presented in Appendices A.1 and A.2.

H. Bagheri, M. Schmitt, and X. X. Zhu. **Fusion of TanDEM-X and Cartosat-1 elevation data supported by neural network-predicted weight maps**. In: ISPRS Journal of Photogrammetry and Remote Sensing 144 (2018), pp. 285–297 [108].

H. Bagheri, M. Schmitt, and X. X. Zhu. **Fusion of Urban TanDEM-X Raw DEMs Using Variational Models**. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11.12 (2018), pp. 4761–4774 [109].

3.1.1 Multi-sensor DEM Fusion Supported by Neural Network-predicted Weights

As illustrated in Section 1.4.2, WA is frequently used for DEM fusion purposes because of its simple implementation and low computational cost. In addition, most advanced techniques apply weights to assist the fusion process to reach the desired output. This means the weights play a key role for efficient fusion of DEMs, especially in the case of multi-sensor DEM fusion, like stereoscopic-optical and InSAR DEMs [9]. The critical problem with using DEM fusion approaches, especially for WA, is applying appropriate weight maps receptive to each DEM—which used to be proportional to the expected height residuals. For this purpose, prior knowledge about existing DEM errors will always be beneficial for the fusion process. One solution for predicting the expected errors is based on an error propagation analysis through the DEM generation procedure. However, usually, such a model can only be an approximation and may not model all potential error sources. An alternative is to learn the error patterns by comparing exemplary areas of interest and corresponding ground truth reference data: e.g., derived from high-precision LiDAR measurements. In this way, suitable weights can be predicted for newly incoming datasets for which neither detailed information about the height errors nor any ground truth data are available. In this regard, a sophisticated framework for appropriate weight map prediction is proposed.

Figure 3.1 displays the framework of the proposed DEM fusion algorithm. In the heart of the proposed framework, an ANN is used to learn the relationship patterns of height errors and corresponding DEM features, which can subsequently be used for forecasting weight maps. The proposed framework (Figure 3.1) can be summarized in three main steps:

1. spatial feature extraction from DEMs and height error calculation
2. data refinement
3. a) training the ANN on dedicated training subsets for which ground truth data is available and b) applying the ANN parameters to target subsets.

The output of the ANN is a predictive model that works as a weight predictor in target areas to which DEMs are fused based on the patterns explored in training subsets. More details of the framework's steps will be explained in the following.

For the training of the ANN, training data are selected from representatives of different land types that can usually be observed over urban areas. From those, different kinds of spatial features describing landscaping and roughness properties of the land surface such as slope, aspect, Anisotropic Coefficient of Variation (ACV), Topographic Ruggedness Index (TRI), Topographic Position Index (TPI), roughness, ruggedness, Surface Roughness Factor (SRF), entropy, edginess, and HEM are extracted. Several studies clarify the relationship between the spatial features and DEM qualities [29, 110, 111, 112]. Figure 3.2 exemplarily shows the maps for these features extracted from the Cartosat-1 data in the industrial area. Moreover, height residual maps are calculated for all training subsets by subtracting the LiDAR ground truth elevations from the corresponding DEM elevations.

Prior to constituting the ANN structure, another important step is to refine the height errors oriented to extracted spatial feature values to get rid of outliers and decrease the noise effects. The calculated height residuals are polluted by high-frequency noise, which will affect the training of the ANN. The performance of the network in the case of using smoothed residual maps derived from the refinement step, as well as using raw data without smoothing, only removing the outliers, are illustrated in Figure 3.3. Figure 3.3(b) indicates that noisy height residuals disrupt the training procedure and prevent the ANN from recognizing the error patterns. Without implementing the refinement, the training performances of networks

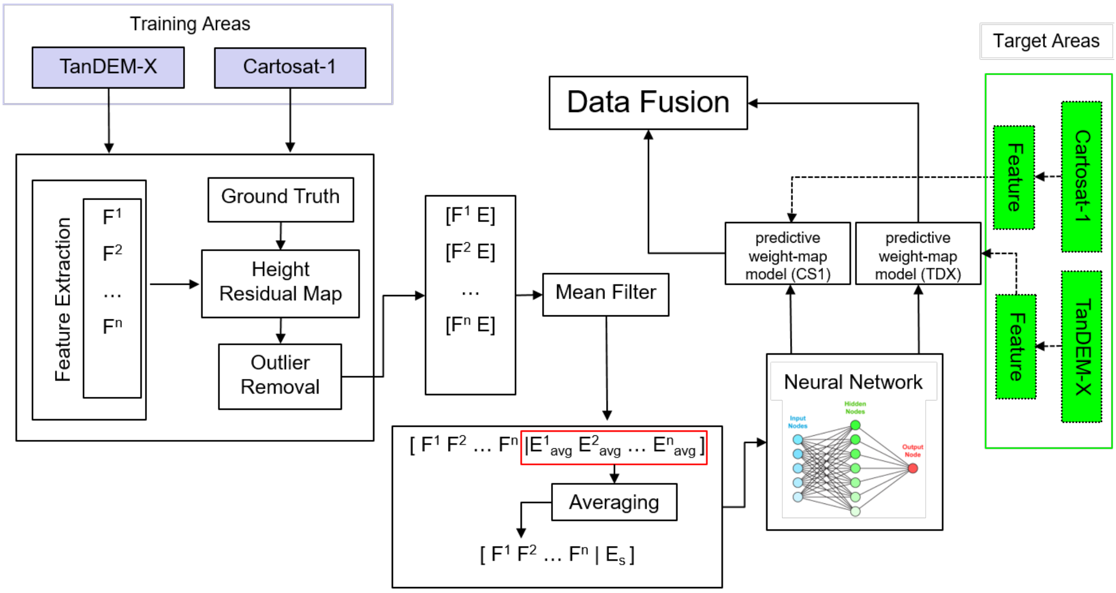


Figure 3.1: A framework designed to estimate the adaptive weights by ANN for the TanDEM-X and Cartosat-1 DEM fusion

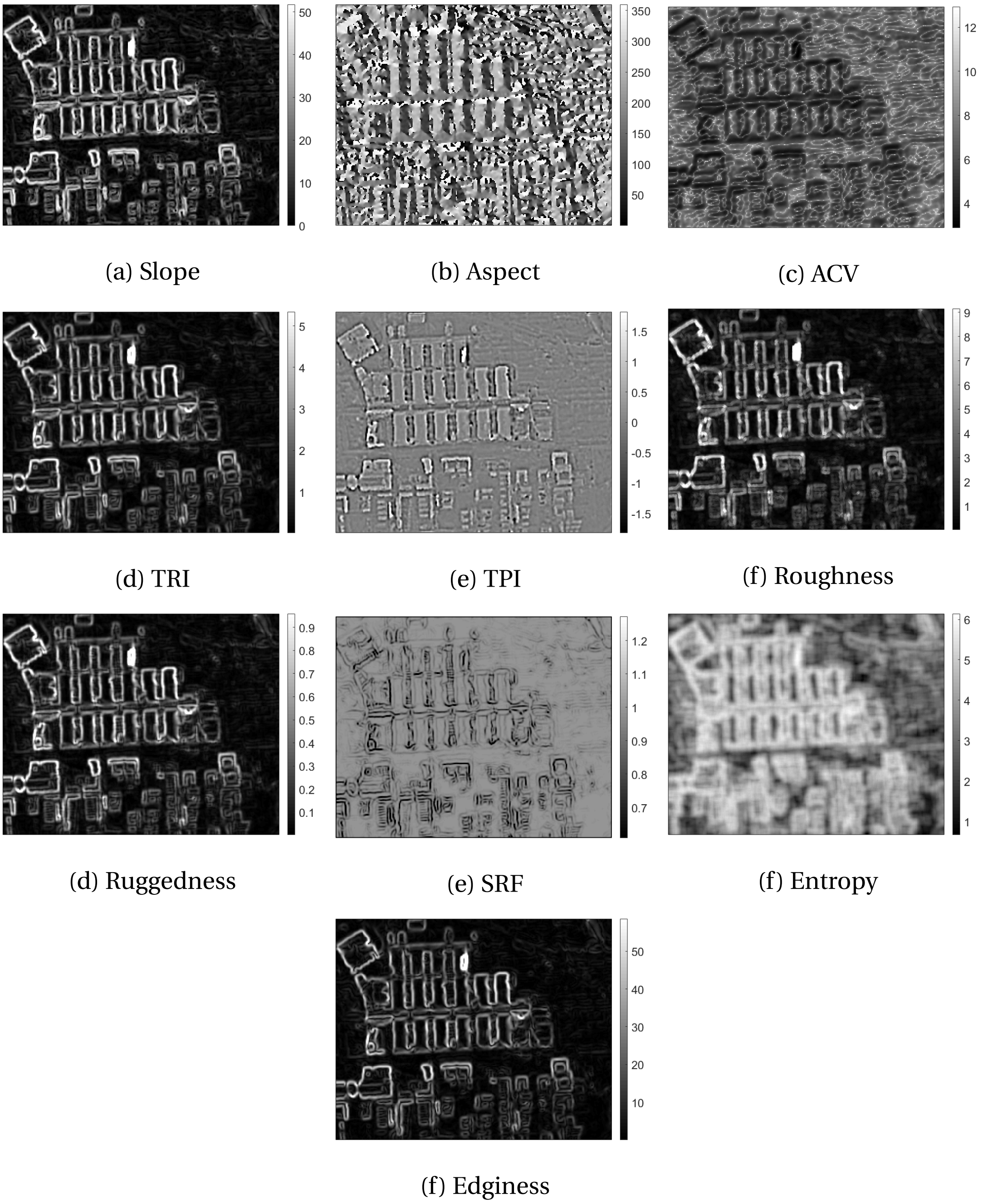
are significantly lower, exemplarily illustrated by output with correlations lower than 0.50 and 0.35 for the TanDEM-X DEM and the Cartosat-1 DEM, respectively.

To reduce the noise effects with the aim of promoting the training, a smoothing process, characterized by two-step mean filtering is carried out. The first step of the refinement is to bin the feature values that can be obtained by a simple empirical-statistical binning technique. At first, errors exceeding 3×NMAD are detected as outliers and then eliminated along with their corresponding feature values from the training dataset. After removing outliers, the values of the feature vector j ($\mathbf{F^j}$) and their corresponding height residuals **E** are binned by the Freedman-Diaconis rule [113]:

$$N = \frac{f^j_{max} - f^j_{min}}{h}, \tag{3.1}$$

where $h = 2 \times I \times k^{-1/3}$. The output of above formulation is the number of bins (N) for feature j with bin width of h, just by detecting the max and min values of measured feature (f^j_{max} and f^j_{min}). I is the interquartile range and k is the number of measurements that are remaining height residuals after outlier removal. In other words, k refers to number of pixels whose height errors (e_i) are lower than the threshold 3×NMAD. The mean filter is applied bin-wise to generate smoother height residual. The output of this filtering is the numerical feature-error model in which each feature vector $\mathbf{F^j}$ corresponds to a new smoothed height residual map, $\mathbf{E^j_{avg}} = \begin{bmatrix} e^j_{1avg} & e^j_{2avg} & \dots & e^j_{mavg} \end{bmatrix}^T$. It has to be noted that infrequent feature values are thrown away by a threshold. This procedure should be followed for each type of feature.

Figure 3.4 presents the graphical depictions of feature-error models derived for an



(a) Slope (b) Aspect (c) ACV

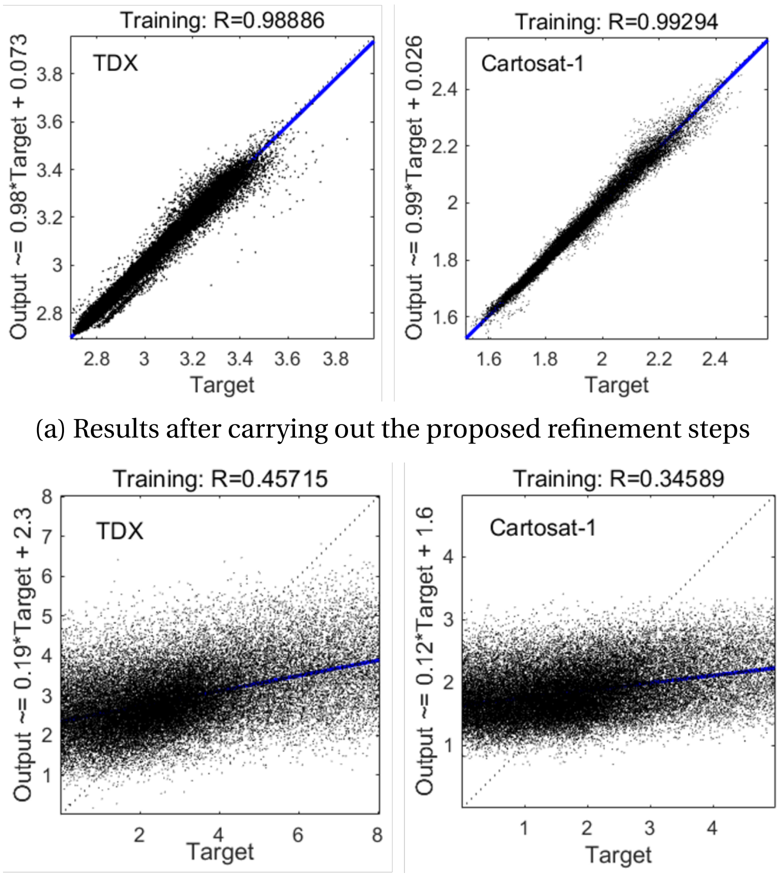
(d) TRI (e) TPI (f) Roughness

(d) Ruggedness (e) SRF (f) Entropy

(f) Edginess

Figure 3.2: Feature maps extracted from DEM from industrial area (D1) of Cartosat-1 DEM

industrial area for TanDEM-X and Cartosat-1 after binning and mean filtering. Consequently, for each pixel, n height residual values at last are acquired. This means there are n residual maps, which are linked to n numerical feature-error models



(a) Results after carrying out the proposed refinement steps

(b) Results without using the proposed refinement

Figure 3.3: Regression plot of training data for industrial area (subset D1)

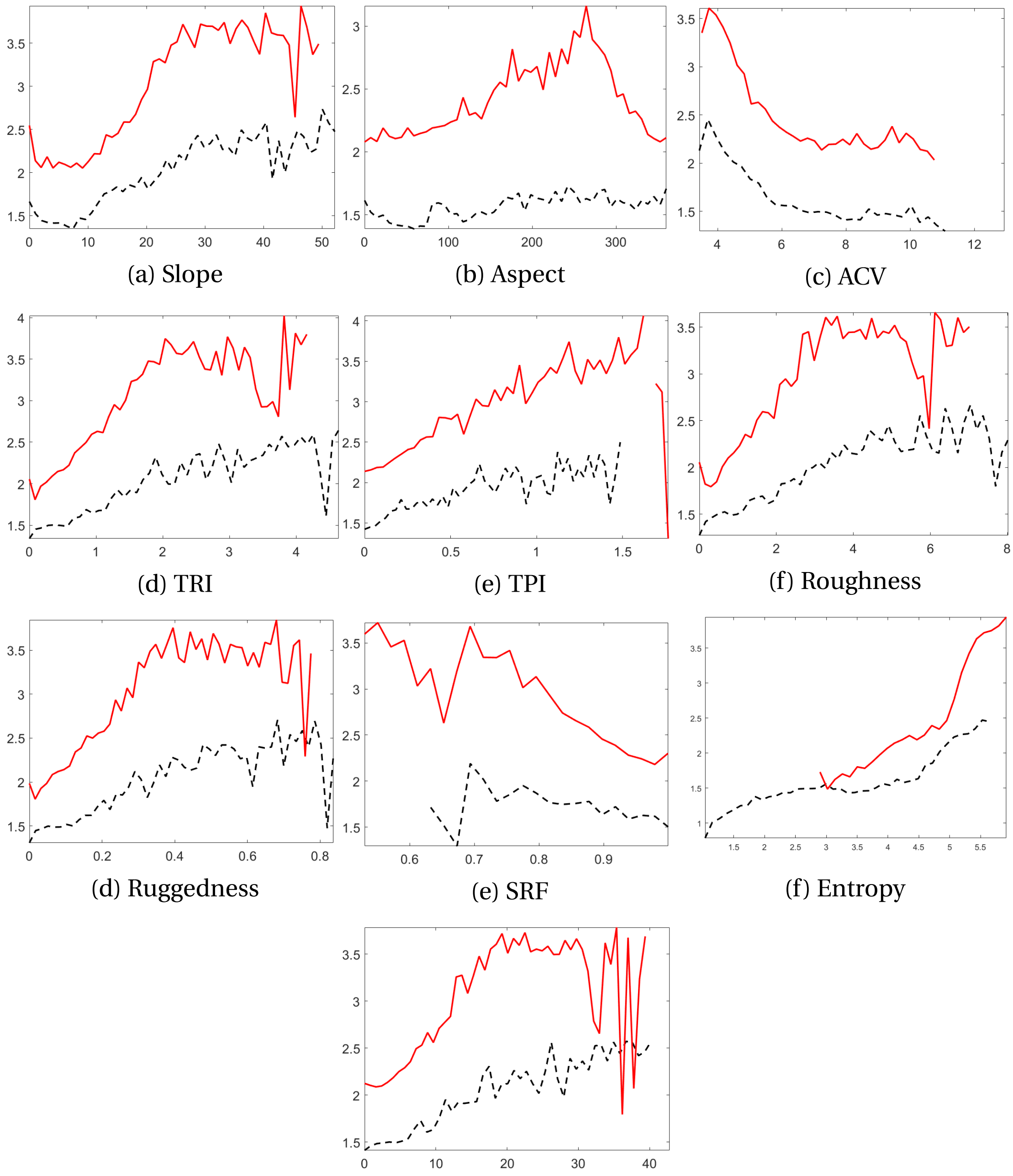
$([\ \mathbf{F^1}\quad \mathbf{F^2}\quad \dots\quad \mathbf{F^n}\quad |\quad \mathbf{E^1_{avg}}\quad \mathbf{E^2_{avg}}\quad \dots\quad \mathbf{E^n_{avg}}\ ])$.

Next, the second step of the smoothing process is to average again the achievements of the former step (smoothed height residuals) to finally create a unique height residual. After this refinement, the data are ready to insert into the ANN for training and exploring the patterns.

The filtered outcomes from the refinement stage are employed to train a fully connected feed-forward neural network. The ANN is trained using the filtered feature vectors as inputs and the modified height residuals as outputs, which are cast in the form of:

$$\begin{aligned} &[\ \mathbf{\Phi_1}\quad \mathbf{\Phi_2}\quad \dots\quad \mathbf{\Phi_m}\quad |\quad \mathbf{E_s}\ ], \\ \text{where}\quad \mathbf{\Phi_i} = [\ f_i^1\quad f_i^2\quad \dots\quad f_i^n\ ]^T, &\quad i \in \{1, 2, \dots, m\} \end{aligned} \tag{3.2}$$

contains the values of the different features for a given pixel i and E_s is the final smoothed residual map obtained through the two-step mean filtering. Figure 3.5 shows the structure of the network, which consists of an input layer in which neurons with the label of the feature values of each pixel ($\mathbf{\Phi_i}$) are connected to the smoothed height residual of the corresponding



(a) Slope

(b) Aspect

(c) ACV

(d) TRI

(e) TPI

(f) Roughness

(d) Ruggedness

(e) SRF

(f) Entropy

(f) Edginess

Figure 3.4: Height error patterns of the TanDEM-X DEM (in red) and the Cartosat-1 DEM (in black dashed) for the industrial area (subset D1): horizontal directions show the feature values and vertical directions indicate mean absolute height residuals in each bin achieved from the refinement step

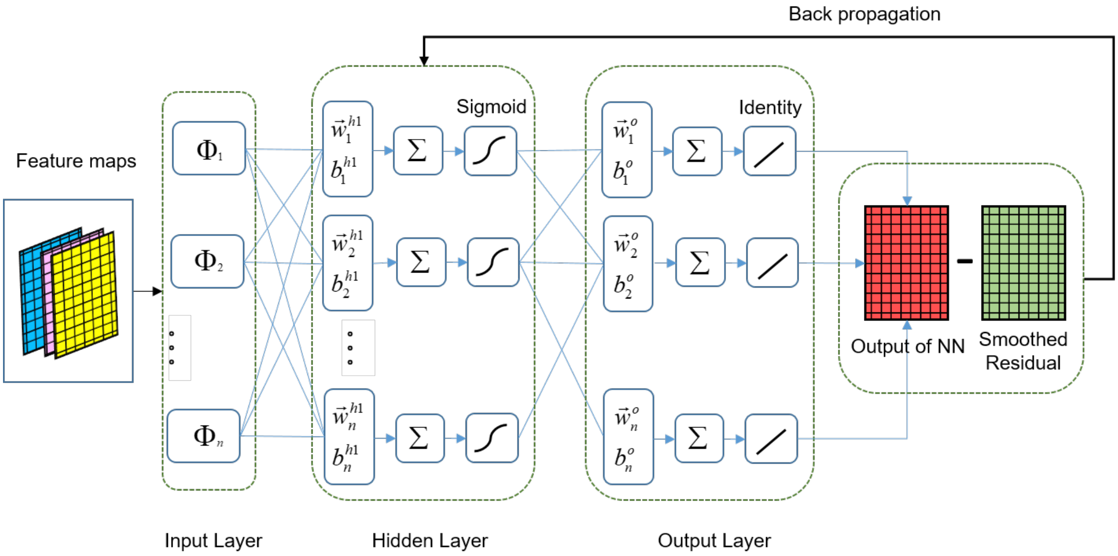

Back propagation
Feature maps
Φ_1
Φ_2
Φ_n
Sigmoid
$\vec{w}_1^{h1}$ b_1^{h1}
$\vec{w}_2^{h1}$ b_2^{h1}
$\vec{w}_n^{h1}$ b_n^{h1}
Σ
Identity
$\vec{w}_1^{o}$ b_1^{o}
$\vec{w}_2^{o}$ b_2^{o}
$\vec{w}_n^{o}$ b_n^{o}
Output of NN
Smoothed Residual
Input Layer
Hidden Layer
Output Layer

Figure 3.5: Structure of neural network for weight map prediction

pixel through the hidden layers. In the repetitive process, with back propagation training, the weights of neurons are gradually modified to decrease the discrepancy between smoothed height residual maps and the map achieved by the network. The main achievement of the ANN after successful training is that a model can estimate the weight maps for each part of DEMs from forecasting the height residuals just by measuring the spatial features.

3.1.2 Multi-modal DEM Fusion Using Variational Models

As explained in Section 3.1, one possibility for improving the quality of medium-resolution DEMs such as TanDEM-X DEM in difficult terrains is fusion with elevation data derived from a different sensor using an ANN-based fusion pipeline. However, the ground truth data used for weight map generation is not necessarily available for every arbitrary study area. Apart from WA, another major part of fusion techniques are based on variational models to decrease noise in DEMs while preserving the sharpness. As illustrated in Section 1.4.2, this type of fusion technique has been mostly tried to fuse VHR DEMs, particularly in urban areas. Thus, this study investigates the potential of using variational models for medium-resolution DEM fusion such as TanDEM-X over urban areas. In comparison to WA, variational models can efficiently fuse DEMs without a requirement to weights. Thus, these models are mostly applied for fusing multi-modal DEMs [30, 31].

Variational models were firstly used for signal and image denoising [114, 115]. A favorite type of variational models is the TV-based model in which the gradient of the desired output image is selected to form the regularization term based on different norms. The main advantage of the TV-based variational model is its convexity that guarantees to find a solution by minimizing the energy functional [31].

In the problem of DEM fusion, several input DEMs are fused using variational models. The data term makes the fused DEM similar to the input tiles while the TV-based regularization term is defined to provide a sharp output at the end by preserving the edges and reducing the noise.

The basic gradient-based variational model for image denoising and data fusion is a quadratic model in which L_2 norm is used for both regularization and data terms [116]. However, the quadratic regularization term causes over-smoothing for edges. Therefore, using the L_1 norm instead was proposed by Rudin, Osher, and Fatemi which is called ROF model correspondingly [114]. Since the ROF model still uses the L_2 norm for the data term, it does not provide robustness against outliers when applied to DEM fusion. As a solution, the L_1 norm can be substituted for the L_2 norm [117]. The TV-L_1 model consists of the data fidelity and the penalty term:

$$\min_{\mathbf{f}} \left\{ \sum_{i=1}^{k} \|\mathbf{f}-\mathbf{h}_i\|_1 + \gamma \|\nabla \mathbf{f}\|_1 \right\}, \tag{3.3}$$

where $\mathbf{h}_i$ are noisy input DEMs and $\mathbf{f}$ is the desired DEM should be achieved by minimizing the functional energy above. The penalty term is formed based on the gradients of the newly estimated DEM to preserve the edges at the end. The regularization parameter γ trades off between penalty and fidelity terms. Increasing γ will influence the smoothness and will produce a smoother fused DEM in the end.

While the main advantage of TV-L_1 is its robustness against strong outliers as well as edge preservation [31], it suffers from the staircasing effect, a phenomenon that creates artificial discontinuities in the final output and particularly affects high resolution DEM fusion [118]. Moreover, the L_1 norm is not necessarily the best choice for all data fusion and denoising cases. As an alternative, the Huber regularization model is proposed to rectify the drawbacks of the TV-L_1 model [31]. It applies the Huber norm instead of the L_1 norm in both fidelity and penalty terms [119]:

$$\|x\|_\eta = \begin{cases} \frac{|x|^2}{2\eta} & \text{if } |x| \leq \eta. \\ |x| - \frac{\eta}{2} & \text{if } |x| > \eta. \end{cases} \tag{3.4}$$

Here, η is a parameter that determines a threshold between the L_1 and L_2 norm in the model. Based on this, the Huber model can be defined as [120]

$$\min_{\mathbf{f}} \left\{ \sum_{i=1}^{k} \sum_{\Omega} \|\mathbf{f}-\mathbf{h}_i\|_\alpha + \gamma \sum_{\Omega} \|\nabla \mathbf{f}\|_\beta \right\}, \tag{3.5}$$

where both data and penalty terms are constituted based on the thresholds α and β that are substituted as η in the Huber norm relation (3.4) to form these terms and Ω denotes the raster DEM space. It should be noted that the Huber norm is a generalized form of the L_1 norm. However, in this study the Huber norm is also used to strictly penalize the outliers. Using quadratic norm in the regularization term penalizes high-frequency changes more than L_1 norm, and thus, it reduces the noise at the cost of over-smoothing edges.

It is mathematically proven that the TV-based energy functional using either L_1 or Huber norm is convex. One popular strategy for finding the minimum of a convex optimization is to reformulate the functional energy as primal-dual problem [121]. More details of the dual-problem algorithm used for minimizing equations (3.3, 3.5) can be found in A.2 and [121].

3.1.3 Experiments

In this dissertation, TanDEM-X is selected as a great source of elevations for urban 3D reconstruction with a potential of large-scale modeling. As mentioned earlier, TanDEM-X DEM is not perfect in urban areas, and DEM fusion techniques can be applied for improving its quality. In the following, summaries of results from two main experiments of the TanDEM-X DEM quality enhancement through the DEM fusion are presented. In the first experiment, the TanDEM-X DEM is fused with Cartosat-1 DEM using the weights generated by the proposed ANN-based framework. In the second experiment, variational models as advanced fusion techniques are employed for TanDEM-X raw DEM fusion instead of WA.

3.1.3.1 TanDEM-X and Cartosat-1 DEM Fusion Using Proposed ANN-based Framework

Before DEM fusion, the height accuracies and errors of both the TanDEM-X and the Cartosat-1 elevation data for different land types are investigated. The output of this assessment can illustrate the performance of each DEM for a specific land type and allows for an educated judgment about which DEM shows favorable accuracy for which land type, thus providing a basis for the consideration of DEM fusion. The main characteristics of the employed TanDEM-X and Cartosat-1 DEMs are presented in Table 3.1.

Raw TanDEM-X DEM		Cartosat-1 DEM	
Center incidence angle	38.25°	Stereoscopic angle	31°
Equator crossing direction	Ascending	Max number of rays	11
Look direction	Right	Min number of rays	2
HoA	45.81m	Horizontal reference	BKG orthophotos*
Total number of looks	22	Vertical reference	SRTM DEM
Pixel spacing	0.2 arcsec	Pixel spacing	5 m
HEM mean	1.33 m	Mean height error (1σ)	2-3 m

Table 3.1: Properties of TanDEM-X and Cartosat-1 tiles. * For more information look up

For height accuracy assessment as well as DEM fusion experiment, study subsets representing six different land types that are usually found over urban areas and their surroundings are chosen. The subsets are selected from Industrial areas (D1, D2), Inner-city areas (I1, I2), High buildings (H1, H2), Residential areas (R1, R2), Forested areas (F1 ,F2), classical land farms (F1, A, L) and Lake (L) which are displayed in Figure 3.6. All these test areas are located in the area of Munich, Bavaria, so that for each of them a highly-accurate LiDAR point cloud (with a density of at least 1 point per m^2), provided by the Bavarian surveying administration, is available as a reference.

Before uncertainty assessment and subsequently implementing the fusion process, the Cartosat-1 and TanDEM-X data should be homogenized in terms of horizontal and vertical

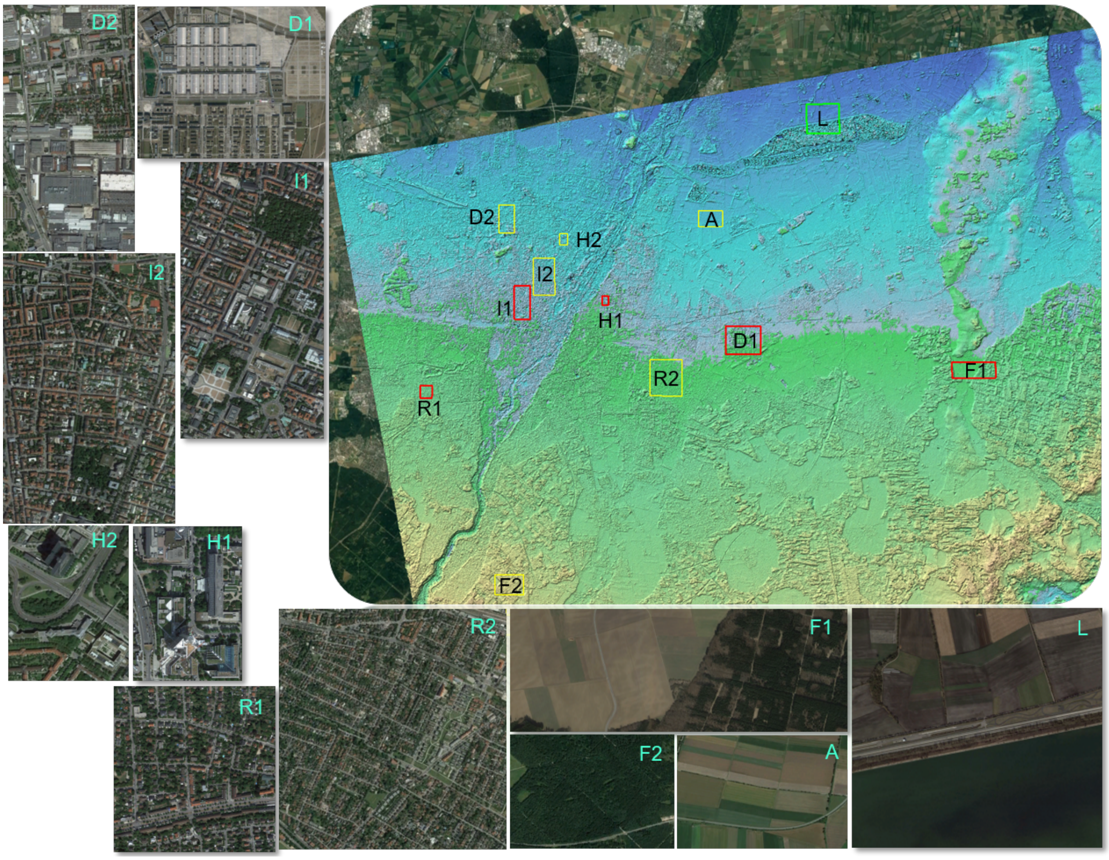
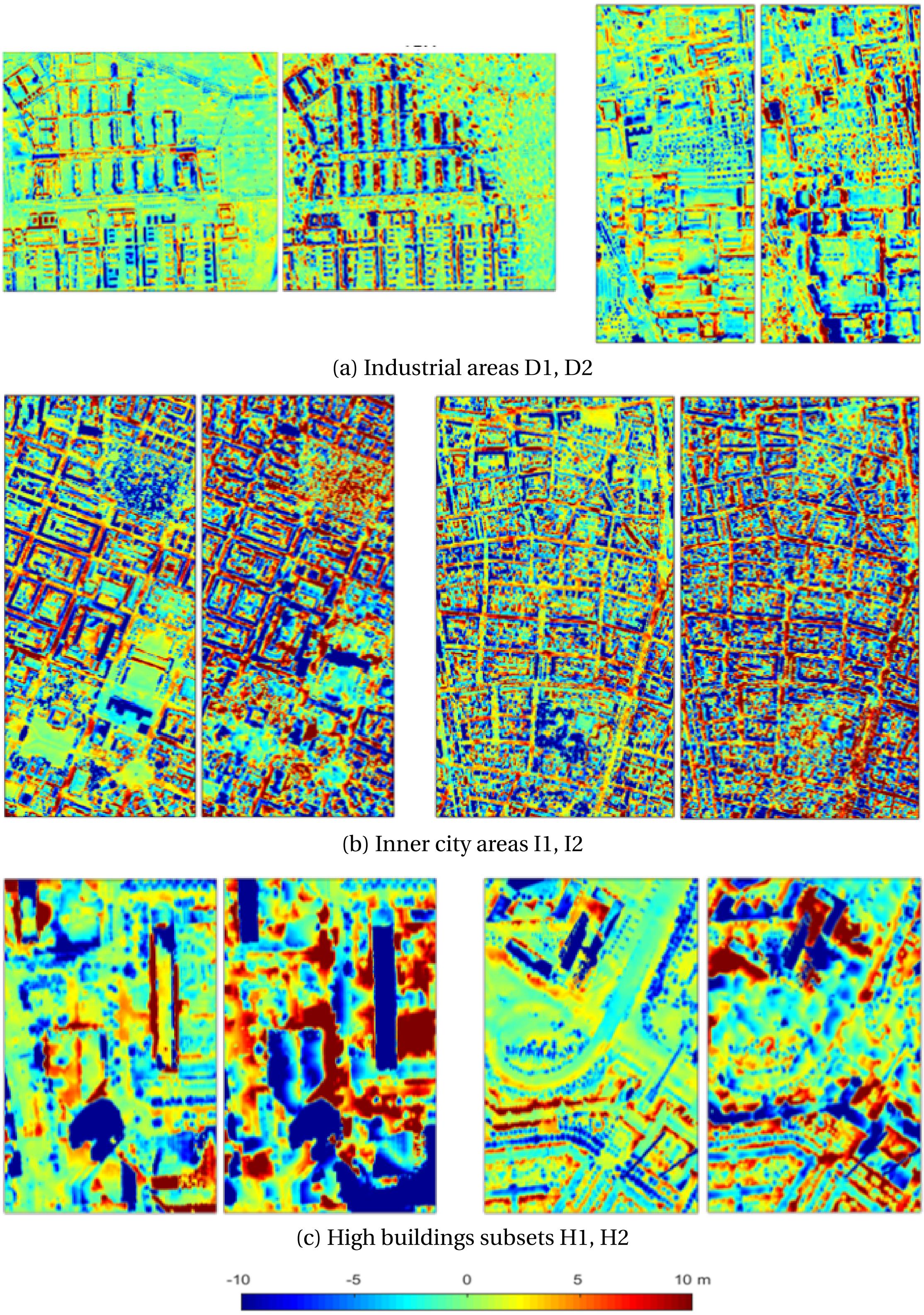


Figure 3.6: Display of land types and locations of study subsets from Munich area

references, as well as pixel spacing. More technical details were already explained in Section 2.3.3. In addition, after preparing the elevation data in the form of the desired grid and reference, the tiles of study DEMs (TanDEM-X and Cartosat-1) should be aligned together to compensate any rotational and translational discrepancies.

At first, the vertical offset of the TanDEM-X and the Cartosat-1 DEM in comparison to the reference DEM in each subset was calculated. On average, the vertical offset of the Cartosat-1 elevation data over the different test areas is 2.657 m while the mean absolute height offset of the TanDEM-X raw DEM over all subsets is found to be 1.503 m. The precision assessment results over urban and non-urban areas after vertical bias removal are collected in Table 3.2. The height residual maps for some exemplary urban study subsets extracted from the TanDEM-X and the Cartosat-1 DEM are displayed in Figure 3.7.

The results of the precision assessment of the TanDEM-X and Cartosat-1 DEM collected in Table 3.2 show that the Cartosat-1 DEM has a higher height precision than the TanDEM-X DEM in urban areas. The main source of errors in the TanDEM-X DEM comes from the layover effect which leads to wrong height reconstructions [122]. In contrast, the TanDEM-X and the Cartosat-1 DEMs have almost identical height accuracy in agricultural and forested



(a) Industrial areas D1, D2

(b) Inner city areas I1, I2

(c) High buildings subsets H1, H2

Figure 3.7: Height residual maps of subsets located in different land types (In each residual map pair, left: Cartosat-1, right: TanDEM-X

Areas		TanDEM-X			Cartosat-1		
		LE90 (STD)	LE90 (NMAD)	RMSE	LE90 (STD)	LE90 (NMAD)	RMSE
Urban	Industrial D1:	7.18	3.88	4.38	**5.41**	**2.69**	**3.29**
	Industrial D2:	8.33	4.98	5.09	**5.67**	**3.85**	**3.46**
	Inner city I1:	11.20	9.28	6.82	**8.31**	**6.56**	**5.11**
	Inner city I2:	9.79	9.11	5.96	**8.06**	**6.88**	**4.94**
	High building H1:	29.22	10.63	18.37	**20.23**	**5.07**	**12.44**
	High building H2:	13.83	6.04	8.44	**11.96**	**3.65**	**7.33**
	Residential R1:	5.09	4.49	3.10	**4.16**	**3.65**	**2.55**
	Residential R2:	4.10	3.34	2.51	3.75	2.97	2.32
Non-Urban	Forested F1:	7.48	7.30	4.56	7.19	6.864	4.371
	Forested F2:	7.30	5.86	4.47	7.55	5.287	4.60
	Agricultural F1	2.31	1.84	1.43	3.23	1.95	1.98
	Agricultural A	1.38	1.03	0.84	1.28	1.06	0.78
	Agricultural L	2.70	1.64	1.64	2.71	1.29	1.65
	Lake (L)	21.42	17.00	14.58	**2.93**	**2.64**	**1.87**

Table 3.2: Height precision (in meter) of TanDEM-X and Cartosat-1 DEMs over different areas. The bold values indicate the best metric values for the respective land type if the margin between the TanDEM-X and Cartosat-1 data is at least 10%

areas.

By successful alignment of the Cartosat-1 and TanDEM-X DEMs, the Cartosat-1 DEM is vertically positioned in the location of TanDEM-X, which indicates an improvement of the absolute geolocation through DEM fusion. After vertical alignment, the ANN-based approach is examined to increase the height precision of both DEMs. The training data are selected from diverse land types introduced earlier. The usage of training data from different land types guarantees the presence of all possible values of features respective to height residuals in the process of pattern recognition by the NN, and give the assurance of discovering a more general model that can be used for any arbitrary land type. After successful training, the ANN can be applied for predicting the height residual in selected target areas where two DEMs are supposed to be fused. The predicted residual maps are used as weight maps in the WA fusion. Thus, two separate ANNs are needed for both the TanDEM-X and Cartosat-1 DEM to generate individual weight maps for each DEM separately.

All subsets of all different land types are simultaneously used as training data to create a general predictor model that can be used for weight map generation in arbitrary target subsets. In this experiment, the subsets D1, I1, H1, R1, and F1 are used to train the ANN, and the resulting output model is used for all target subsets. 70% of data from the training subsets are devoted to training, and 15% are for validation to control the training process in order to avoid over-fitting and under-fitting, tuning the networks' parameters such as depth and number of neurons in each layer, and the remaining data (15%) are devoted to monitoring the performance of NN during the training. However, the whole process of the proposed framework will be implemented on the independent subsets (D2, I2, H2, R2 and F2, A) for measuring the performance of the DEM fusion pipeline.

Table 3.3 compares the results of ANN-based fusion with simple HEM-based fusion. The ANNs with one hidden layer and 20 neurons in the hidden layer were applied for this DEM fusion.

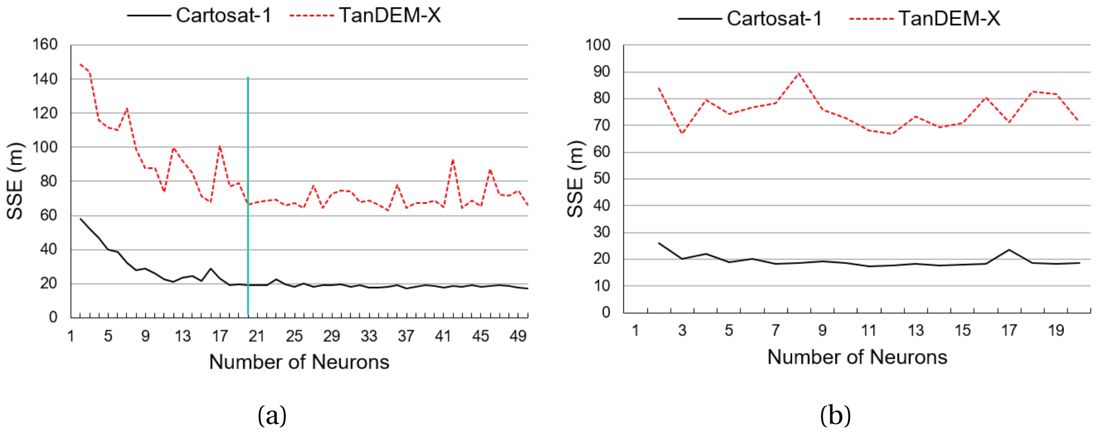
Areas	TanDEM-X		Cartosat-1		Fused DEM			
					HEM		ANN	
	NMAD	RMSE	NMAD	RMSE	NMAD	RMSE	NMAD	RMSE
Industrial: D2	3.43	5.11	2.43	3.56	2.91	4.26	**2.42**	**3.46**
Inner city: I2	5.92	5.81	**4.62**	5.21	5.09	5.13	4.90	**4.84**
High building: H2	3.75	8.41	3.29	8.13	3.14	7.93	**3.13**	**7.75**
Residential: R2	2.30	2.61	2.02	2.45	2.13	2.44	**1.99**	**2.33**
Forested: F2	3.88	4.82	3.23	4.65	3.51	4.35	**3.19**	**4.18**
Agricultural: A	0.57	0.84	0.65	0.81	0.98	0.76	**0.49**	**0.68**

Table 3.3: Results of fusion of TanDEM-X and Cartosat-1 DEM (in meters) by using weight maps generated by different methods. The bold values indicate the best values of metrics relevant to the quality of the compared DEMs

The optimal structures of ANNs for both study DEMs were investigated by tracking the cost function values during the training stage. The performances of networks were evaluated by considering one hidden layer and changing the number of neurons in this layer. Then, deeper networks were examined by adding another hidden layer. Figure 3.8 depicts the performance of the neural networks on test data for an increasing number of neurons in the ANNs with one hidden layer, as well as for structures designed with two hidden layers. The plots display that the rise in the number of neurons in the first layer is more influential than adding the layers and making the networks deeper. In ANNs with one hidden layer, the general trend of changing costs remained stable after adding more than 20 neurons.

Figure 3.9 visualizes the absolute residual maps of TanDEM-X, Cartosat-1, HEM, and ANN-based DEM fusion results in comparison to reference data for some exemplary study areas. As the obtained NMAD and RMSE results show, standard HEMs generally cannot be reliably used to produce a fused DEM whose height precision exceeds the Cartosat-1 DEM precision. This confirms the assumption that standard HEMs do not reflect all possible error sources in the original DEM data. As an example, the HEM delivered with the TanDEM-X raw DEM just contains error values derived from interferometric coherence and baseline configuration, while deterministic error sources such as layover are not considered. Nevertheless, standard HEMs can be used as a fall-back solution should ground truth for ANN-based weight map prediction be unavailable.

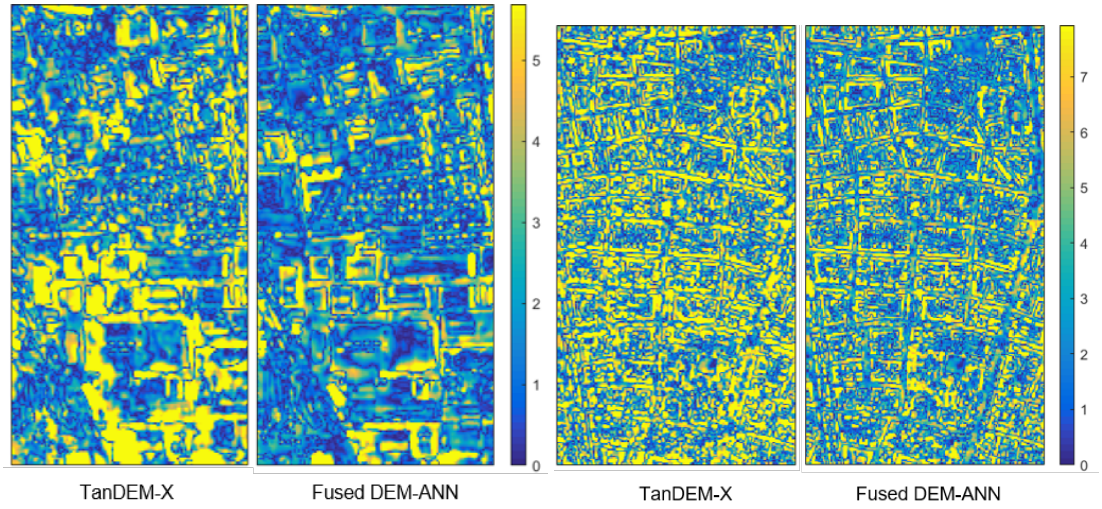
In contrast, the results obtained for the ANN-supported fusion shows an improvement of the fused DEM product with respect to both input datasets, indicating that the designed ANNs can properly model the existing error patterns related to spatial features that describe the landscaping and the roughness of the land surface under investigation. While the ANN-supported DEM fusion significantly improves the height precision of the TanDEM-X DEM, it also enhances the quality of the Cartosat-1 DEM: As an additional analysis reveals, more than 51% of all fused DEM pixels are more accurate than their Cartosat-1 counterparts.



(a)

(b)

Figure 3.8: The performance of the ANNs with different structures measured by SSE (Sum of Squared Errors); a) The structure with one hidden layer. b) The structure that organized by two hidden layers: first, with number of neurons fixed to n=20, and second with varying number of neurons



Subset D2

Subset I2

Figure 3.9: Absolute residual maps of TanDEM-X and fused DEM using ANN-predicted weight maps in some exemplary subsets

Last but not least, the absolute vertical accuracy of the fused DEM is also better than the absolute accuracy of the original Cartosat-1 DEM, which is achieved through the alignment to the more accurately localized TanDEM-X DEM. Thus, eventually, the proposed DEM fusion can lead to a final DEM product that provides a higher quality than the individual input DEMs in both absolute and relative measures.

3.1.3.2 TanDEM-X Raw DEM Fusion by TV-based Variational Models

A current approach for implementing the TanDEM-X raw DEM fusion in DMP is WA. After WA-based DEM fusion, visualization shows that outlines of buildings are not perfectly sharp and still some amount of existing noise spoils the footprints of buildings. Another possible solution for the TanDEM-X quality improvement is to use variational models instead of WA in DMP. For this aim, different experiments are carried out to evaluate the performance of variational models such as TV-L_1 and Huber models for multi-modal TanDEM-X DEM fusion.

The first investigation includes data takes that have similar baseline configurations as well as HoAs. The study subsets are selected from two nominal TanDEM-X raw DEMs over Munich city in Germany. The characteristics of these raw DEMs are presented in Table 3.4. From those, four subsets as representatives of different land types are extracted for the DEM fusion task.

TanDEM-X raw DEMs: Munich area			
Acquisition Id	1023491	1145180	1058842
Acquisition mode	Stripmap	Stripmap	Stripmap
Center incidence angle	38.25°	37.03°	38.33 °
Equator crossing direction	Ascending	Ascending	Ascending
Look direction	Right	Right	Right
Polarization	HH	HH	HH
HoA	45.81m	53.21m	72.02m
Pixel spacing	0.2 arcsec	0.2 arcsec	0.2 arcsec
HEM mean	1.33 m	1.58	2.58

Table 3.4: Properties of the TanDEM-X raw DEM tiles for Munich area

The results of raw DEM fusion using TV-L_1 and Huber models for study areas are presented in Table 3.5. In addition to statistical analysis, to evaluate the performance of variational models, the residual maps of the input DEMs and the fused DEMs achieved by different methods for the industrial and inner-city 1 study areas are displayed in Figure 3.10. The results illustrate using variational models in the fusion process can finally improve the quality of the TanDEM-X DEM over the quality achievable with classic WA. It is explicitly displayed on residual maps that variational models can finally reduce the noise effects and also makes the footprints of buildings more apparent than WA.

In the second experiment, the performance of variational models for fusing TanDEM-X raw DEMs with different HoAs over urban areas is investigated. For this purpose, one experimental ITP raw DEM with different HoA over Munich city in Germany is considered. The specifications of this raw DEM are shown in Table 3.4 (tile 1058842).

In this experiment, a study subset is extracted from an area that has lots of inconsistent heights due to PU errors. For this aim, a relatively large subset from an urban area which is covered by trees and also includes a river crossing is selected. Figure 3.11 displays the selected study area suffering from PU errors. The corresponding DEM data are derived from tiles 1023491 and 1058842 with HoAs about 45 m and 72 m respectively.

The PU errors appearing in this subset originate from the volume decorrelation phe-

Study area	DEM		Mean	RMSE	MAE	NMAD	STD
Industrial	Raw DEM	id: 1023491	0.71	4.40	3.08	2.37	4.34
		id: 1145180	0.71	4.64	3.27	3.01	4.58
	Fused DEM	WA	0.77	4.16	2.93	2.24	4.09
		TV-L_1	**0.69**	**3.67**	**2.69**	**2.03**	**3.60**
		Huber	0.71	3.74	2.84	2.40	3.67
Inner 1	Raw DEM	id:1023491	0.78	7.79	5.95	6.49	7.75
		id:1145180	0.78	8.08	6.30	7.15	8.04
	Fused DEM	WA	0.84	7.51	5.83	6.49	7.46
		TV-L_1	**0.77**	**6.11**	**5.00**	5.72	**6.06**
		Huber	0.78	6.14	5.09	**5.67**	6.09
Inner 2	Raw DEM	id:1023491	0.18	7.00	5.44	6.36	7.00
		id:1145180	0.18	7.16	5.57	6.51	7.16
	Fused DEM	WA	0.20	6.82	5.33	6.23	6.82
		TV-L_1	**0.12**	5.83	**4.78**	**6.16**	5.83
		Huber	0.18	**5.82**	4.82	6.23	**5.82**
Residential	Raw DEM	id:1023491	0.95	2.68	2.10	2.05	2.50
		id:1145180	0.95	2.92	2.25	2.31	2.76
	Fused DEM	WA	0.96	2.61	2.05	1.99	2.43
		TV-L_1	**0.89**	**2.41**	**1.96**	**1.98**	**2.24**
		Huber	0.95	2.44	1.98	**1.98**	**2.24**
Agricultural	Raw DEM	id:1023491	0.13	0.86	0.57	0.59	0.84
		id:1145180	0.13	1.64	1.13	1.20	1.64
	Fused DEM	WA	0.14	0.78	0.51	0.54	0.76
		TV-L_1	**0.06**	**0.55**	**0.29**	**0.20**	**0.54**
		Huber	0.13	0.72	0.48	0.47	0.71
Forested	Raw DEM	id:1023491	**2.25**	4.84	3.54	3.46	4.28
		id:1145180	**2.25**	4.58	3.36	3.24	3.99
	Fused DEM	WA	2.28	4.51	3.30	3.17	3.89
		TV-L_1	**2.25**	**4.34**	**3.18**	**3.09**	**3.71**
		Huber	**2.25**	4.36	3.21	3.12	3.73

Table 3.5: Height accuracy (in meter) of the TanDEM-X data before and after DEM fusion in the different study areas over Munich

nomenon that happens in an area covered by trees (like the selected study subset) and also a coherence change due to transition from dry land to water (river). PU errors typically are at the range of multiples of the HoA value. The inconsistent heights can be determined by [26]

$$dh_{th} = 0.75 \times min(|HoA|) - 4. \tag{3.6}$$

Those height residuals bigger than dh_{th} are denoted as inconsistent height values emerging because of PU errors.

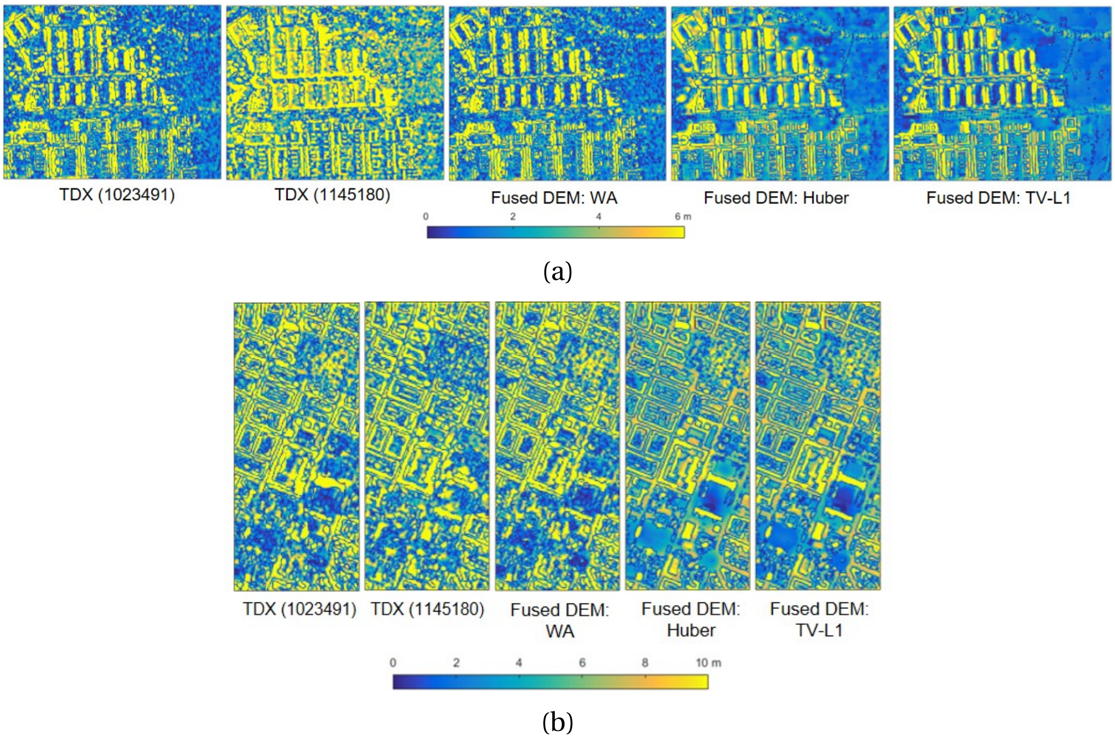


Figure 3.10: Absolute residual maps of the initial input raw DEMs and the fused DEMs obtained by different approaches for the industrial (a) and inner city (b) study areas over Munich

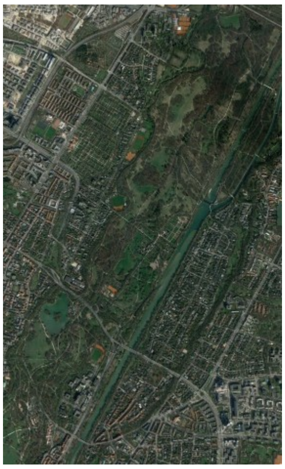


Figure 3.11: The study subset selected for DEM fusion in a problematic area (4.5 km × 2.8 km)

Table 3.6 collects the results of fusing DEMs with different HoAs in the selected study area. Moreover, Table 3.7 compares the fused DEMs with different approaches and initial DEMs in terms of the number of PU errors, maximum and minimum height residuals. The PU threshold for each DEM is computed based on the respective HoA value using (3.6). It is evident that the DEM 1058842 has the lower number of PU errors because of larger HoA, but for DEM fusion quality analysis, the minimum value of HoAs (here 45.81) is considered to enumerate the number of PU errors.

DEM		Mean	RMSE	MAE	NMAD	STD
Raw DEM	id: 1023491	**-2.35**	10.77	8.46	10.10	10.51
	id: 1058842	**-2.35**	10.57	8.27	9.69	10.30
Fused DEM	WA	-2.37	10.45	8.23	9.81	10.17
	TV-L_1	-2.63	9.24	7.13	8.03	8.86
	Huber	**-2.35**	**8.60**	**6.70**	**7.65**	**8.277**

Table 3.6: Height accuracy (in meter) of the TanDEM-X data with different HoAs before and after DEM fusion in the problematic study area

DEM		HoA	PU Threshold	No. of PU Errors	Max Discrepancy	Min Discrepancy
Raw DEM	id: 1023491	45.81	30.36	2032	51.80	-73.13
	id: 1058842	72.02	50.01	51	58.82	-54.76
Fused DEM	WA	45.81	30.36	1339	50.74	-53.39
	TV-L_1	45.81	30.36	102	19.16	-33.76
	Huber	**45.81**	**30.36**	**0**	**16.97**	**-28.71**

Table 3.7: Effect of DEM fusion to reduce the number of PU errors using tiles with different HoAs in the problematic study area

The results from Tables 3.6 and 3.7 demonstrate the efficiency of the Huber model for the fusion of two tiles of TanDEM-X raw DEMs in the problematic area. This proves, in addition to DEM fusion methodology, selecting appropriate raw DEM tiles dependent on the problem is significant for a successful fusion.

In the final experiment, it is focused on the fusion of DEMs acquired by different baseline configurations including different orbit directions and HoAs. Table 3.8 provides the properties of the tiles used for this experiment.

The raw DEMs covering Terrassa and Vacarisses cities located in Spain were produced by ascending and descending acquisitions. In addition to orbit directions, the HoAs of tiles is also not similar to each other. Again, from these tiles, study subsets located in different land types were selected.

The results of fusing ascending and descending raw DEMs in different land types over the urban area are provided in Table 3.9. The results of DEM fusion again illustrate using variational models can increase the accuracy of the initial input raw DEMs. In urban study subsets, the performance of the Huber model is slightly better than TV-L_1 according to the statistical metrics, but their differences are not really significant. It can be concluded that both models produce similar results in terms of statistical measurements. In comparison to WA, variational models also give a more accurate DEM in urban areas and the agricultural subset.

In conclusion, The results of all experiments illustrated that variational models are more

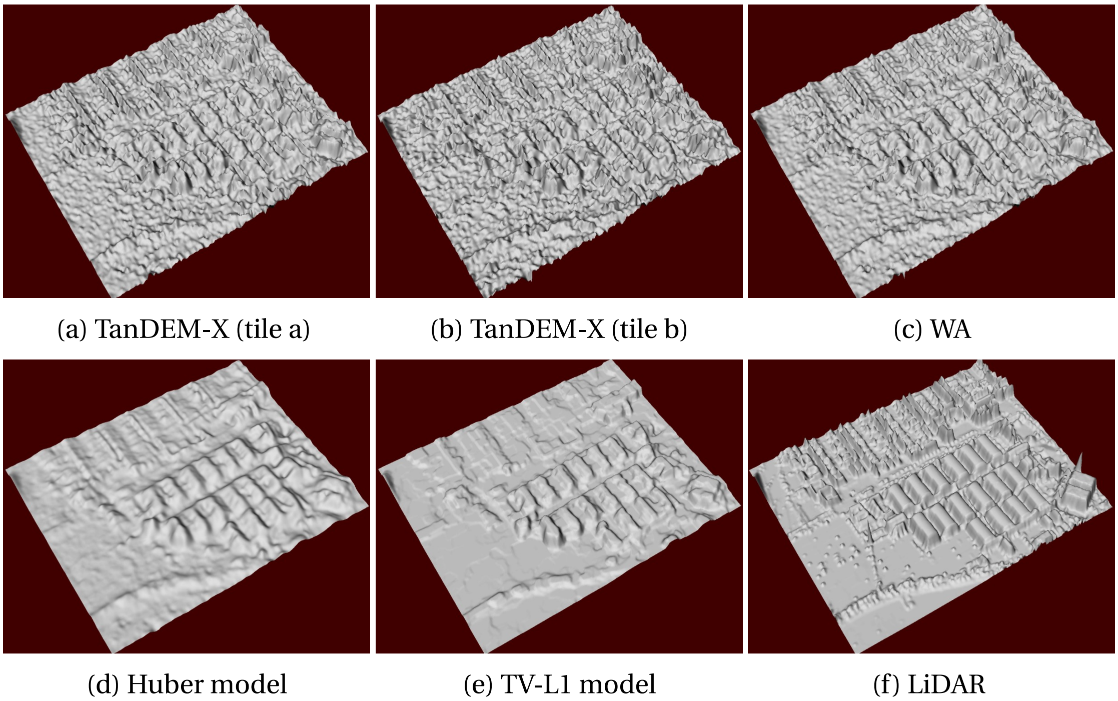
	TanDEM-X raw DEMs	
Acquisition Id	1058683	1171358
Acquisition mode	Stripmap	Stripmap
Center incidence angle	33.71°	34.82°
Equator crossing direction	Ascending	Descending
Look direction	Right	Right
Polarization	HH	HH
HoA	60.18 m	48.58 m
Pixel spacing	0.2 arcsec	0.2 arcsec
HEM mean	1.17 m	1.40

Table 3.8: Properties of the nominal ascending and descending TanDEM-X raw DEM tiles over Terrassa and Vacarisses cities

Study area	DEM		Mean	RMSE	MAE	NMAD	STD
Industrial	Raw DEM	1058683	-0.19	3.49	2.49	2.60	3.48
		1171358	-0.19	3.56	2.44	2.34	3.55
	Fused DEM	WA	-0.26	3.06	2.13	2.07	3.05
		TV-L_1	-0.34	2.92	**2.09**	**2.07**	2.90
		Huber	**-0.19**	**2.89**	2.10	2.14	**2.88**
Inner	Raw DEM	1058683	-0.78	5.05	3.52	3.70	4.99
		1171358	-0.78	5.11	3.53	3.62	5.05
	Fused DEM	WA	-0.76	4.66	3.22	**3.36**	4.59
		TV-L_1	-0.91	4.35	**3.08**	3.40	**4.25**
		Huber	**-0.78**	**4.34**	3.13	3.52	4.27
Residential	Raw DEM	1058683	-0.54	4.24	3.11	3.19	4.20
		1171358	-0.54	4.42	3.21	3.26	4.38
	Fused DEM	WA	-0.62	3.94	2.87	2.83	3.90
		TV-L_1	-0.76	3.96	2.88	2.77	3.88
		Huber	**-0.54**	**3.86**	**2.86**	**2.74**	**3.82**
Agricultural	Raw DEM	1058683	0.44	2.38	1.68	1.71	2.34
		1171358	0.44	1.93	1.23	0.98	1.88
	Fused DEM	WA	0.35	**1.60**	**1.04**	0.83	1.57
		TV-L_1	**0.27**	**1.60**	**1.04**	**0.78**	1.59
		Huber	0.44	1.62	1.12	0.91	**1.56**

Table 3.9: Height accuracy (in meter) of the ascending and descending TanDEM-X data before and after DEM fusion in the different study areas over Vacarisses and Terrassa

efficient than WA for TanDEM-X raw DEM fusion, especially over urban areas. However, the Huber model tends to provide a smoother fused DEM than TV-L1. Figure 3.12 shows that the Huber model produces a smoother output in comparison to TV-L_1 because of mixing the quadratic norm and the L_1 norm to form data and regularization terms.



(a) TanDEM-X (tile a)

(b) TanDEM-X (tile b)

(c) WA

(d) Huber model

(e) TV-L1 model

(f) LiDAR

Figure 3.12: 3D display of initial TanDEM-X raw data and the results of DEM fusions using different methods in the industrial area used in the first experiment.

3.2 Height Generation by SAR-Optical Stereogrammetry

In addition to DEMs which are typically produced by SAR interferometry or optical stereoscopy, height information for 3D reconstruction and building modeling can be provided from archive-stored VHR SAR and optical imagery through a cooperative fusion. This cooperative integration can rectify optical imagery with modern SAR sensors such as TerraSAR-X and consequently provide 3D spatial information with high absolute geolocation accuracy. A review on the literature of 3D reconstruction from SAR-optical imagery identifies that only a limited number of studies have dealt with stereogrammetric 3D reconstruction from high-resolution SAR-optical imagery over urban areas. Moreover, their pipelines for 3D reconstruction are just founded using sensor geometry in a sparse matching manner. Thus, this section presents a summary of investigation on the possibility and potential of implementing a dense multi-sensor stereo pipeline for 3D reconstruction of urban areas from high-resolution SAR-optical image pairs. More details are provided in Appendix A.3.

H. Bagheri, M. Schmitt, P. d'Angelo, and X. X. Zhu. **A framework for SAR-optical stereogrammetry over urban areas**. In: ISPRS Journal of Photogrammetry and Remote Sensing 146 (2018), pp. 389–408 [123].

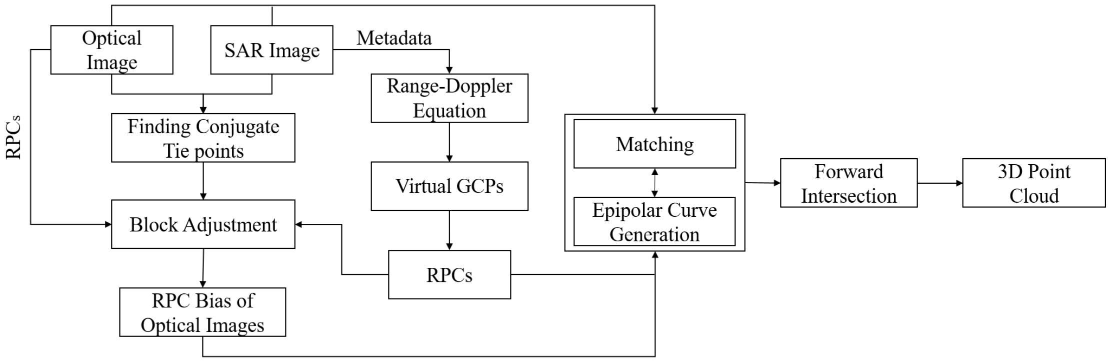


Figure 3.13: Framework for stereogrammetric 3D reconstruction from SAR-optical image pairs

3.2.1 A Framework for SAR-Optical Stereogrammetry

Figure 3.13 shows the general framework of SAR-optical stereogrammetric 3D reconstruction. Similar to optical stereogrammetry, one grayscale optical image and one amplitude SAR image form a stereo image pair that can be processed by suitable matching methods to find all possible conjugate pixels. However, some important pre-processing steps are required before the matching and 3D reconstruction.

Currently, most VHR optical images are delivered using RPCs. Thus, the primary step in the SAR-optical stereogrammetry framework is to estimate the RPCs for SAR imagery as well. This process homogenizes the geometry models of both sensors and simplifies the subsequent processes of SAR-optical block adjustment and establishing an epipolarity constraint. The RPCs are estimated from Virtual GCPS (VGCPs) achieved by the range-Doppler model of SAR imagery. The accuracy of the RPCs can be estimated using independent virtual checkpoints that are produced in a similar way to VGCPs using the range-Doppler model. The accuracies of the fitted RPCs for the TerraSAR-X data acquired over Munich and Berlin study areas are listed in Table 3.10. The analysis was performed based on the residuals of the rows and columns, given by the differences between image coordinates computed by RPCs and range-Doppler. The analysis results confirm that the RPCs can model the range-Doppler geometry for TerraSAR-X data to within a millimeter, and can thus well be used in the 3D reconstruction process.

Table 3.10: Accuracy (STD) of RPCs fitted on SAR sensor model (units: m)

Area	Virtual GCPs		Check points	
	row	column	row	column
Munich	0.00026	0.00114	0.00025	0.00031
Berlin	0.00024	0.00027	0.00026	0.00118

Conceptually, the epipolar plane can not be formed for SAR-optical image pairs due to specific imaging geometries and consequently, the classic, straight forward epipolar geometry does not exist for a SAR-optical image pair. However, using the epipolar curve equation

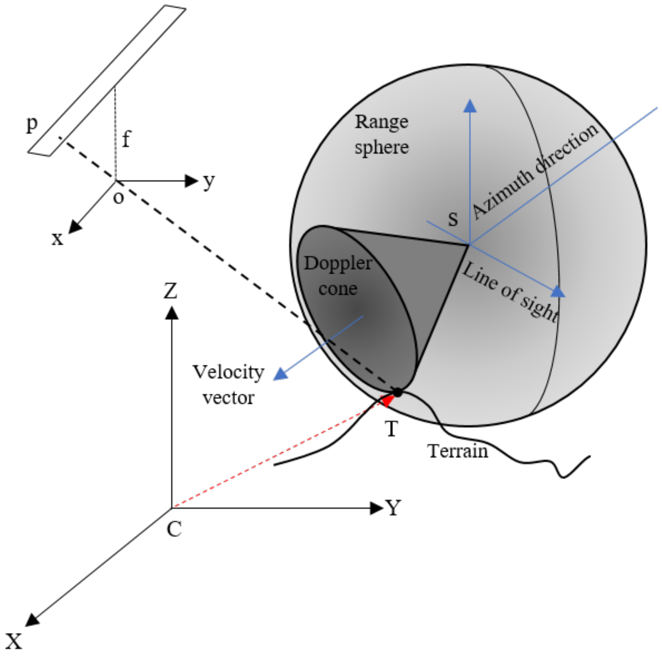


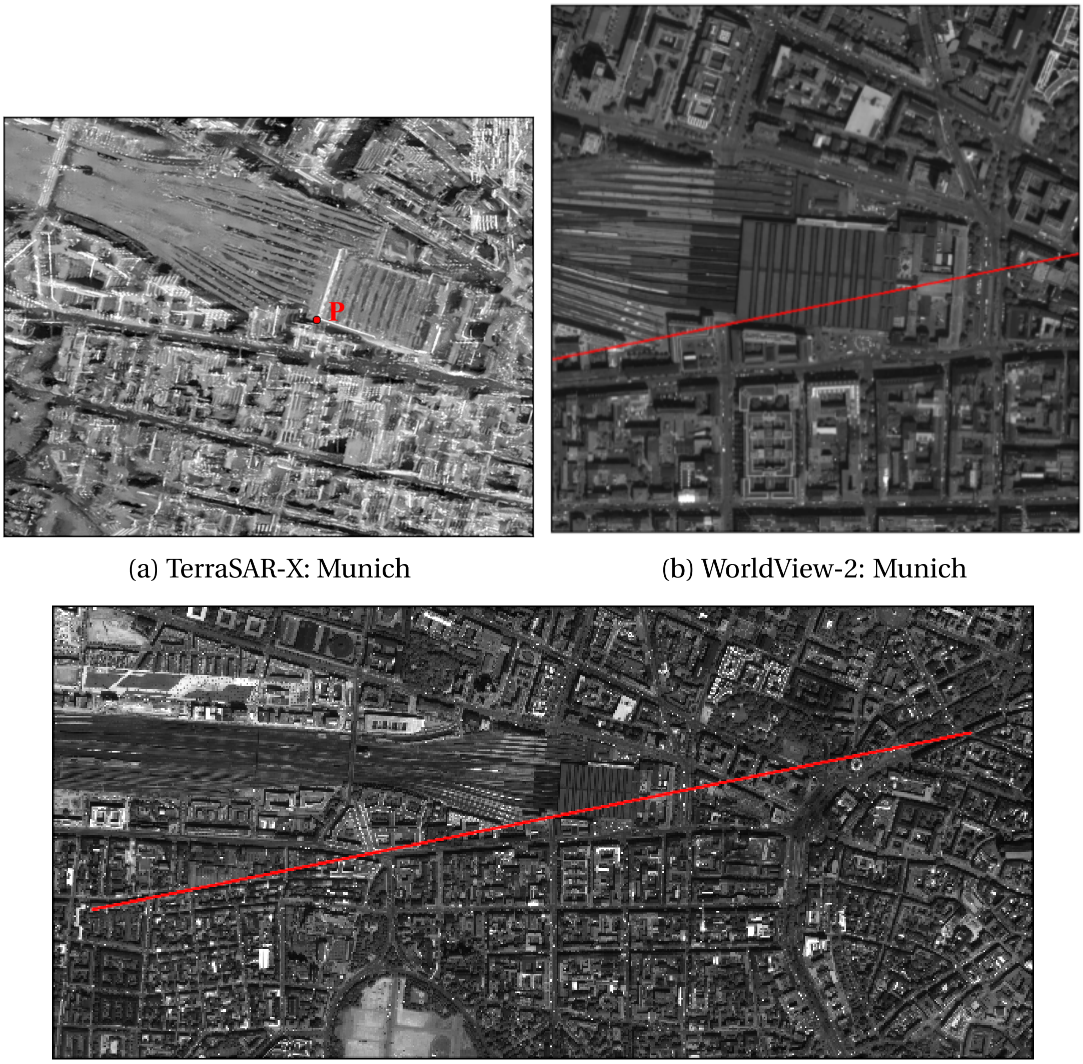
Figure 3.14: Imaging geometry for configuration of SAR-optical imagery

presented in equation (2.5) and the rigorous models describing sensor geometries of SAR and optical imagery (Section 2.5.5), a rigorous model representing epipolarity constraint for SAR-optical image pairs acquired by space-borne platforms can be established. For this task, the optical image is considered as the left-hand image and the SAR image is the right-hand image. Figure 3.14 shows the configuration of the SAR-optical stereo case. Mathematically, it is proved that a general rigorous model representing the epipolarity constraint for SAR-optical image pairs is formulated as (more details in A.3)

$$\Gamma\, y_r = \sqrt{F_2\, x_r^2 + F_1\, x_r + F_0} - R_0. \tag{3.7}$$

This shows that an epipolarity-like constraint can be established for SAR-optical image pairs. However, the non-linear relation between y_r and x_r in equation (3.7) shows that SAR-optical epipolar curves are not straight, even under the assumption of linear motion for the SAR system.

In addition to mathematical proof, the epipolarity constraint will be experimentally investigated for an RPC-based imaging model. The validity of the derived SAR-optical epipolarity constraint is analyzed for an exemplary point located at the corner of the Munich central train station building (p). This point was projected to the terrain space by changing the heights in specific steps, e.g., 10 m, starting from the lowest possible height and proceeding to the highest possible height in the scene (for this experiment, the interval [0 m, 1200 m] is used). The output will be an ensemble of points with different heights, such as depicted in Figure 3.15(c). All these points were then back-projected to the WorldView-2 image space



(a) TerraSAR-X: Munich

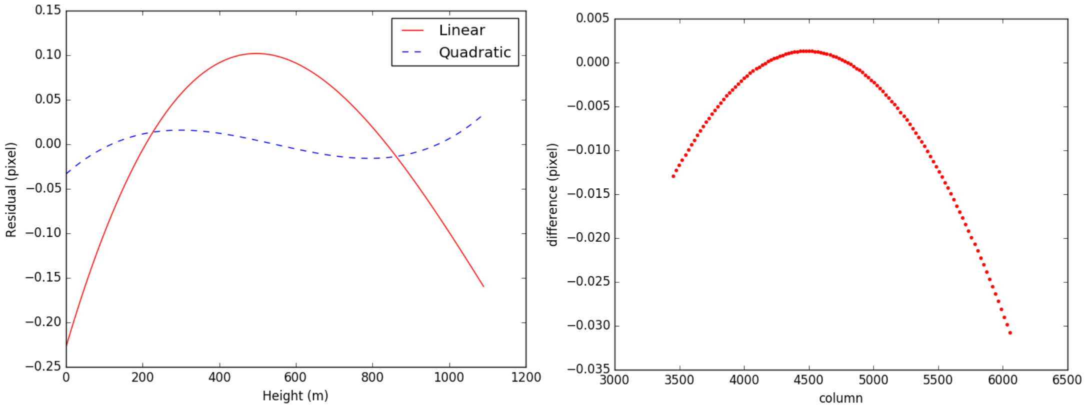
(b) WorldView-2: Munich

(c) WorldView-2: Munich

Figure 3.15: Epipolar curves for the WorldView-2 image (of Munich) given by changing the heights of point *p* (located at the corner of the Munich central train station in the TerraSAR-X scene) for all possible height values in the image scene. The epipolar curves look like straight, but are not actually straight

using RPCs. The corresponding epipolar curve for all possible heights in the study area is constructed by connecting the image points obtained in this way, as shown in Figure 3.15(c). Although the epipolar curve appears to be straight, more analysis is required to determine whether this is the case. By expanding the image, it can be seen in Figure 3.15(b) that the epipolar curve nearly passes through the conjugate point of *p* in the WorldView-2 image.

To clarify the straightness of the epipolar curve constructed for point *p*, linear and quadratic polynomials were fitted to the image points of the epipolar curve. Figure 3.16(a)



(a) (b)

Figure 3.16: a) Linear and quadratic polynomials fitted on the epipolar curves in the WorldView-2 images; b) Difference of two corresponding epipolar curves over the column direction. The maximum difference between the two epipolar curves is less than one pixel

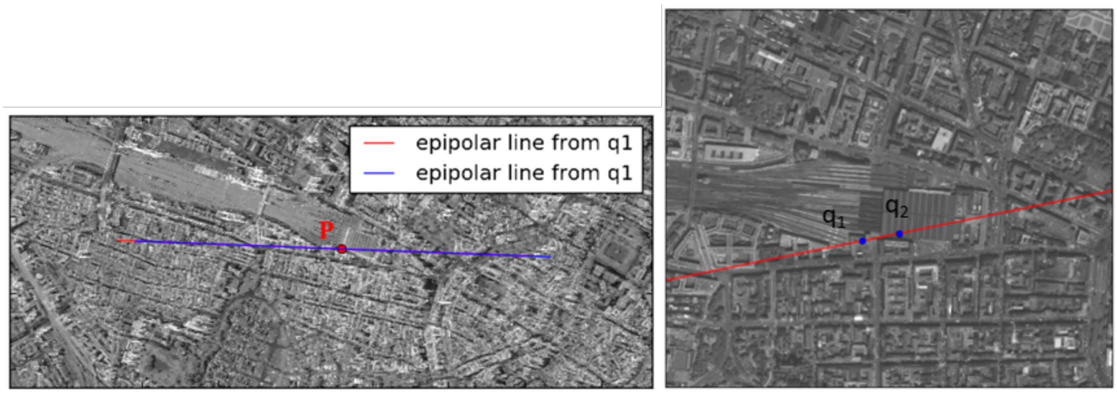


Figure 3.17: Corresponding epipolar curves in the Munich TerraSAR-X image (left) derived from two points, q_1 and q_2 on the epipolar curve of the Munich WorldView-2 image (right)

represents the least-squares residuals with respect to the point heights for the epipolar curves created in both study subsets. The residuals of the linear fit for the epipolar curve established for the Munich WorldView-2 image range from -0.25 to 0.1 pixels (i.e., meters), whereas the residuals of the quadratic fit are close to zero.

To investigate the conjugacy of the SAR-optical epipolar curves, two distinct points (q_1,q_2) were selected from the epipolar curve in the WorldView-2 image. From each of these points, the corresponding epipolar curves were constructed in the TerraSAR-X image for all possible heights as in the experiment before. Figure 3.17 displays the corresponding epipolar curves in TerraSAR-X given by q_1 and q_2 located in the WorldView-2 image. The epipolar curves appear to pass through point p located in the SAR image. Further analysis clarifies that the differences in the column direction between the two epipolar curves passing through point

p are less than one pixel, allowing the matching of the two epipolar curves (Figure 3.16(b)).

Similarly, all experiments as mentioned above were also performed for the TerraSAR-X/WorldView-2 image pair of Berlin and analyses illustrated that the epipolarity constraint could also be established for this image pair. Further information is presented in [123].

The next phase is to carry out a multi-sensor block adjustment to align the optical image to the SAR image. The main peculiarity of multi-sensor block adjustment is to rectify the RPCs of the optical imagery with respect to the SAR imagery without any requirement of GCPs which leads to improving the absolute geolocalization of the optical imagery and correcting the positions of the epipolar curves on the optical imagery . For this aim, tie points, the common points between the SAR and optical images, are obtained by manual or automatic sparse matching between two images. Then the geographic coordinates of the tie points in the SAR image are calculated by the inverse rational functions (equation 2.20) computed for the SAR imagery and the mean height of the study area. The output is a collection of GCPs that can be applied for the RPC rectification of the optical imagery. The resulting GCPs are then projected by the rational function associated with the optical images to give the image coordinates of the GCPs. The block adjustment equations can then be written as

$$F_x^i = -x_o^i + c_{ou}^i + \Delta x^i + v_x^i = 0, \tag{3.8}$$

and

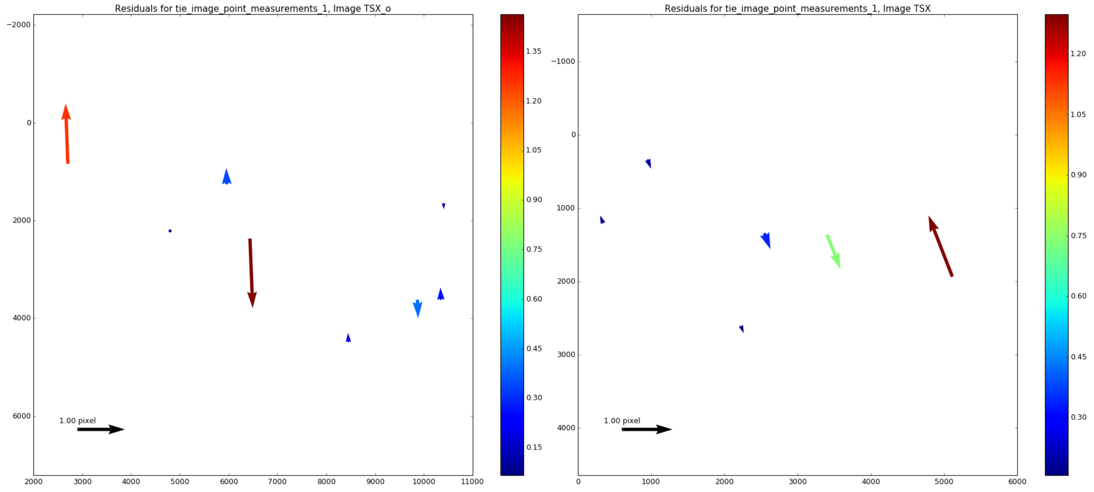
$$F_y^i = -y_o^i + r_{ou}^i + \Delta y^i + v_y^i = 0, \tag{3.9}$$

where, x_o^i and y_o^i denote the column and row of tie point i in the optical scene, and c_{ou}^i and r_{ou}^i are the un-normalized coordinates of the tie point after projection and back-projection using the RPCs. Δx^i and Δy^i are affine corrections presented in equation 2.6. Finally, through an iterative least-squares adjustment [89], the unknown parameters m_i and n_i are estimated and the affine model can be formed. This affine model is added to the rational functions of the optical image to improve the geolocation accuracy to that of the SAR image.

For the experiment, eight and six tie points were selected to match the WorldView-2 images to the TerraSAR-X images in the Munich and Berlin study areas, respectively. Figures 3.18(a) and 3.18(b) show the residuals of the full multi-sensor block adjustment for each tie point. The results demonstrate that the residuals of most points are less than one pixel in both experiments, which indicates a successful implementation of SAR-optical block adjustment. Table 3.11 presents the bias of the row and column components resulting from the block adjustment of WorldView-2 and TerraSAR-X image pairs for both study areas. Figures

Table 3.11: Block adjustment results (units: m)

Area	Sensor	Bias Coefficients		STD	MAD	Min	Max	No. of Tie Points
Munich	WorldView-2	-2.47	-0.53	0.50	0.14	0.07	1.46	8
	TerraSAR-X	0	0	0.50	0.14	0.07	1.46	
Berlin	WorldView-2	-0.73	0.28	0.51	0.13	0.19	1.59	6
	TerraSAR-X	0	0	0.42	0.11	0.16	1.30	



(a) Munich

(a) Berlin

Figure 3.18: Residuals of tie points after full multi-sensor block adjustment in the TerraSAR-X image space.

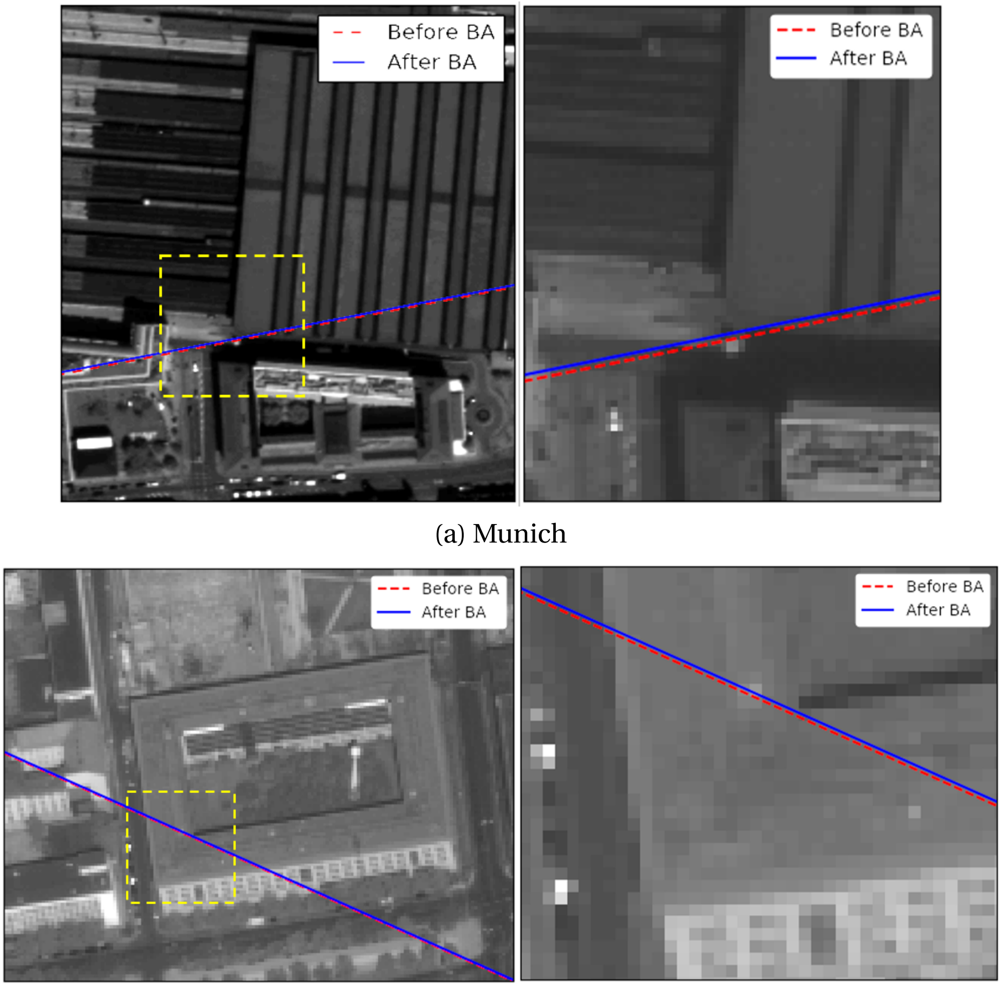
3.19(a) and 3.19(b) display the locations of the epipolar curves before and after adjustment. By enlarging the images, it is possible to confirm that the displacement of the epipolar curves is minimal, yet noticeable.

In the following, the ability to perform dense image matching for SAR-optical image pairs using SGM is investigated. The RPC model is used to realize the SAR-optical epipolar geometry implicitly without the need to generate normal images. In addition, the minimum and maximum disparity values should be selected to restrict the length of the search space along the epipolar curves. These values can be determined using external data such as the SRTM DEM [124]. Among the different matching measures, MI is recommended for SGM as it is known to perform well for images with complicated illumination relationship, such as SAR-optical image pairs [125]. The next setting is to switch off the *minimum region size* option. Experimental results show that, for SAR-optical image pairs, the complex illumination relationship between the images and the different imaging effects (especially for urban areas) make the *minimum region size* criterion useless, as connectivity cannot be ensured in the disparity map. Similar to other dense matching cases, the LR (Left-Right) check is used to investigate binocular half-occlusions [126]. Finally, SGM is implemented at four hierarchy levels, and the aggregated cost is calculated along 16 directions around each point.

A disparity map is then produced in the frame of the reference image via SGM. This disparity map should be transferred from the reference sensor geometry to a terrestrial reference coordinate system such as UTM.

3.2.2 Experiments

For the experiments, two study areas were selected, one in Berlin and one in Munich (both located in Germany). First, the SAR images were filtered with a non-local speckle filter. Then,



(a) Munich

(b) Berlin

Figure 3.19: Displacement of epipolar curves after block adjustment by RPCs. Left images show the epipolar curve positions before and after the bundle adjustment, and right images display the selected patch (identified by dashed yellow rectangles) in an enlarged image

all images were resampled to 1 m × 1 m pixel spacing to enhance the general image similarity and facilitate the matching process. After implementing bundle adjustment for both datasets, two sub-scenes (with a size of 1000 × 1500 pixels each) from overlapped parts of the study areas were cropped. These sub-scenes are displayed in Figure 3.20 .

The dense matching of TerraSAR-X/WorldView-2 imagery produces a sparse point cloud over each of the urban study areas. Among different similarity measures which commonly used in SGM, MI could provide the best results (see A.3). Figures 3.21(a) and 3.21(b) display the reconstructed point clouds from SAR-optical sub-scenes of Munich and Berlin. The difference of SAR and optical observation geometries, the lack of jointly visible scene parts, and the complicated radiometric relationship between SAR and optical imagery cause that stere-

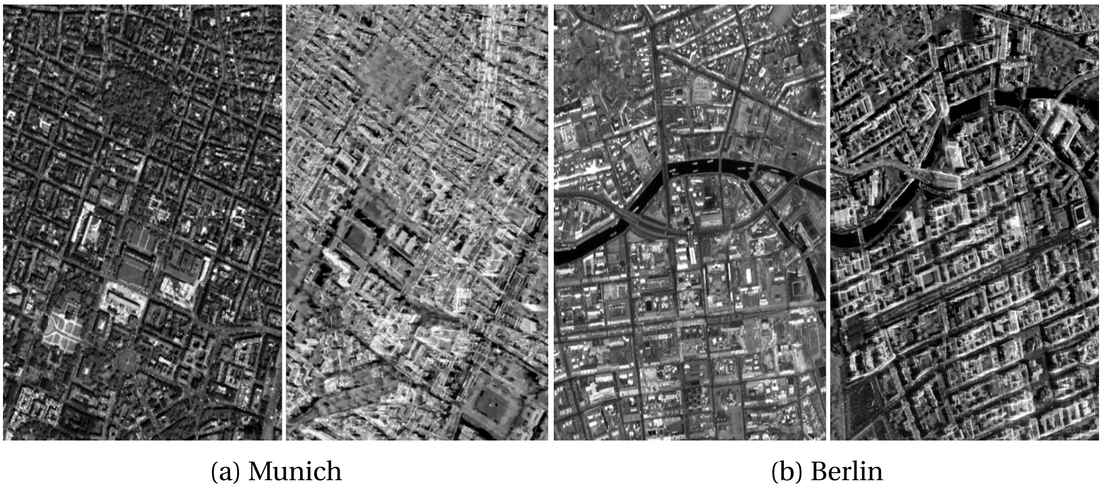


Figure 3.20: Display of SAR-optical sub-scenes extracted from Munich and Berlin study areas. In each pair, the left-hand image is from WorldView-2, right-hand image is from TerraSAR-X

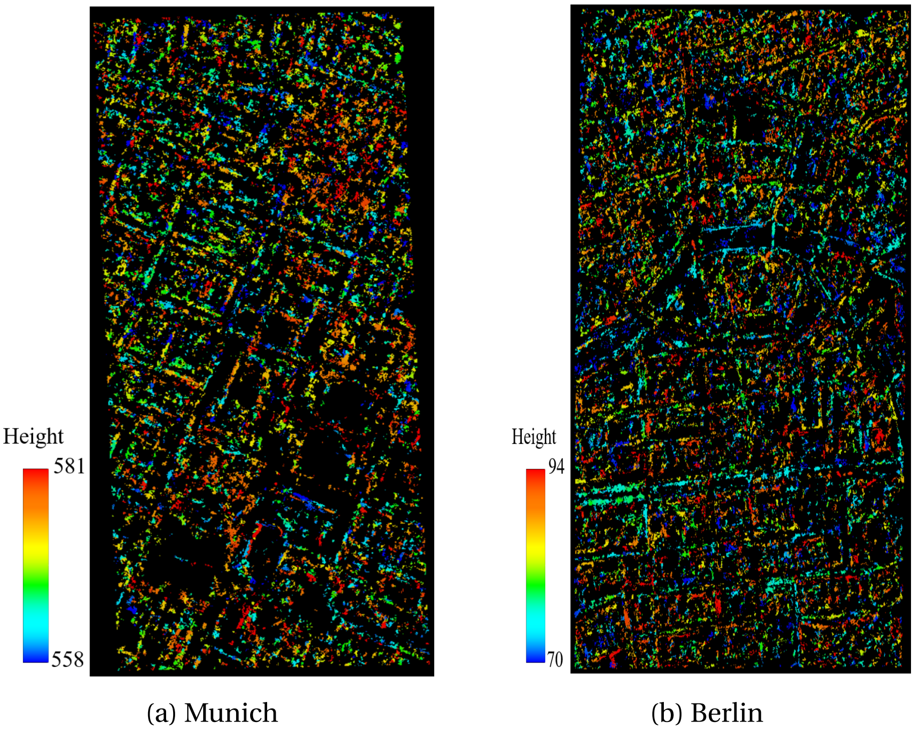


Figure 3.21: Point clouds reconstructed from Munich and Berlin sub-scenes

ogrammetric 3D reconstruction leads to sparse rather than dense point clouds over urban areas.

The accuracy of these sparse point clouds was compared to that of reference LiDAR point clouds along the X, Y, and Z directions by least-squares plane fitting [127]. Table 3.12 summarizes the mean, STD, and RMSE of the distances along the different axes. Table 3.13 additionally shows results corresponding to point clouds that were cleaned by removing points deviating from the SRTM model by more than 5 m.

Table 3.12: Accuracy assessment of reconstructed point clouds with respect to LiDAR reference

Area	Mean (m)			STD (m)			RMSE (m)		
	X	Y	Z	X	Y	Z	X	Y	Z
Munich	-0.003	0.025	0.080	1.285	1.350	2.652	1.285	1.351	2.653
Berlin	0.000	-0.041	0.273	1.566	1.692	3.091	1.566	1.693	3.103

Table 3.13: Accuracy assessment of point clouds after SRTM-based outlier removal

Area	Point Cloud	25%-quantile	50%-quantile	75%-quantile	Mean (m)
Munich	original	0.77	1.89	3.58	2.44
	filtered	0.67	1.56	3.04	2.12
	SRTM	0.73	1.64	3.25	2.21
Berlin	original	0.89	2.01	3.67	2.75
	filtered	0.79	1.76	3.22	2.35
	SRTM	0.86	1.93	3.63	2.65

In conclusion, the quantitative analysis shows that 25% of all points are reconstructed with clear sub-pixel accuracy, while the median accuracy lies at about 1.5 to 2 m. The experiments also show that the results can be further improved by filtering outliers from the reconstructed point clouds. In this study, the globally available SRTM DEM as prior knowledge was employed for outlier removal. As Table 3.13 shows, discarding points with a height difference to SRTM greater than 5m improves the results significantly.

3.3 Urban 3D Reconstruction Using Multi-sensor-derived Heights

The final objective of this dissertation is to use multi-sensor-derived information for LOD1 building modeling of urban scenes. LOD1 models are usually generated automatically by combining building footprints with height values. While high-resolution DEMs or dense LiDAR point clouds are typically used to generate these building models, these height sources are usually not available on a larger scale. Another possibility is to derive elevations from globally available DEM data, but they are often not detailed and accurate enough to provide sufficient input to the modeling of individual buildings. Therefore, this research investigates the possibility of LOD1-based 3D building modeling from both volunteered geographic information and different remote sensing data sources which can potentially be applied on a large scale. More specifically, OSM building footprints and height data derived from dedicated experiments, i.e. DEM fusion and SAR-optical stereogrammetry presented in Sections 3.1, and

3.2 are employed. The following sections provide a summary of the paper in Appendix A.4.

H. Bagheri, M. Schmitt, and X. X. Zhu. **Fusion of Multi-sensor-derived Heights and OSM-derived Building Footprints for Urban 3D Reconstruction**. In: ISPRS International Journal of Geo-Information 8.4 (2019) [128].

3.3.1 Input Data for 3D Building Modeling

In this research, the heights for 3D building reconstruction are provided by different sources. For the experiment, the inner-city subset located in Munich, Germany and commonly used in three previous experiments (Sections 3.1.3, and 3.2.2), was selected. The list of the different input height sources used in this experiment are presented in the following.

- *Cartosat-1 DEM*: The Cartosat-1 DEM used in this study is as same as that applied in Section 3.1.3.1 and its specifications were already presented in Table 3.1.

- *TanDEM-X raw DEM*: Among three tiles of TanDEM-X raw DEM acquired over Munich city, the highly accurate one, tile 1023491 is selected. The properties of this tile are presented in Table 3.4.

- *DEM generated by TanDEM-X and Cartosat-1 DEM fusion*: The output of the TanDEM-X and Cartosat-1 DEM fusion using ANN-based weight map generation pipeline is another source of heights that will be employed in this study. It should be noted through the DEM fusion, the overall quality of the TanDEM-X DEM was improved over urban areas (see Section 3.1.3.1).

- *DEM generated by TanDEM-X raw DEM fusion*: Another experiment for improving the quality of TanDEM-X raw DEMs was implemented in Section 3.1.3.2. Using the variational models for TanDEM-X raw DEM fusion could also lead to a high-quality DEM over urban areas.

- *Point cloud generated by SAR-optical stereogrammetry*: By implementing the SAR-optical stereogrammetry framework for the TerraSAR-X and WorldView-2 image pair, a sparse point cloud could be produced (see Section 3.2). The produced heights can be potentially employed for 3D building modeling.

3.3.2 LOD1 Building Model Generation

For LOD1 reconstruction using OSM-provided footprints, two scenarios will be considered. The first one is to model buildings based on primary footprint layers provided by OSM. The second is to update building outlines in the primary footprint layer of OSM. This updating is implemented because of defects in building footprints of OSM in which a building can consist of several intra-blocks with different heights while the blocks appear as a single outline with the same heights in OSM. Thus, modification of building outlines respective to height changes should be considered. In more details, As displayed in Figure 1.1, for a building consisting of two blocks, each with different height level, it may appear as an integrated building outline in OSM, and thus, only one height value can be assigned for it while the outline should be split into two outlines. In other words, the heights lying in two actually separate clusters are substituted by a median value located somewhere in the middle. While this ultimately leads to a significant height bias, modifying and producing a precise outline map can decrease the risk of bias appearance in the final reconstruction.

In this research, this building modification is performed semi-automatically. The candidate outlines are detected by clustering heights. The number of clusters determines the number of height levels and implies potential building blocks. Then, this is verified by looking up to LiDAR data or even open satellite imagery such as Google Earth. Horizontal displacements of OSMs' building footprints respective to highly accurate data such as LiDAR can also lead to a height bias. This phenomenon makes the inclusion of some non-building points in building outlines, and due to substantial height differences between non-building and building points, the final height estimations are affected by bias. In this study, inside building points are selected by buffering from each outline inwards.

Figure 3.22 displays LOD1 3D reconstruction results for the study area consisting of prismatic building models generated by combining the height information derived from different sources discussed in the previous experiments (Sections 3.1, and 3.2) and building footprints provided by OSM. As displayed in Figure 3.22, all models systematically underestimate the building heights in comparison to a model produced from high-resolution LiDAR data. However, this underestimation becomes minimum for a model using heights derived from SAR-optical stereogrammetry, as can be seen when comparing large buildings. However, for better evaluation, quantitative assessment should be performed. Therefore, the height accuracy of each LOD1 model was validated by comparing it with a model was created from the reference LiDAR DSM in a similar manner. For that purpose, the original LiDAR point cloud first is interpolated to a grid with a 1 m pixel spacing. Then, TV-L_1 denoising [109] is used to reduce potential noise effects. This TV-L_1 denoising mitigates biases in building height estimation induced by height outliers and inconsistencies such as those caused by crane-towers. Then, the final height of each building outline can be computed according to the process described earlier. Those can be correspondingly applied for the quality measurements of the 3D building reconstructions obtaining from other height information sources. The quantitative evaluations for the LOD1 reconstructions implemented based on scenario 1 (using original OSM) and 2 (using updated outlines) are presented in Tables 3.14, and 3.15, respectively.

Regarding the median values in Table 3.14, using the original building outlines causes a bias affecting estimated final heights (RMSE values) while standard deviations are much smaller, thus confirming a systematic change in building heights. This bias can be significantly reduced by modifying building outlines in a preprocessing step (Table 3.15).

The results demonstrate that using heights derived from outputs of multi-sensor DEM fusion can still lead to better reconstruction in comparison to the primary TanDEM-X DEM. While the highest accuracy is obtained by Cartosat-1 data, it owes the accuracy to the bias compensation through the alignment to TanDEM-X. Without the alignment, the existing bias would be propagated to the final building heights.

Last but not least, it has to be mentioned that for generating a complete 3D city model, it

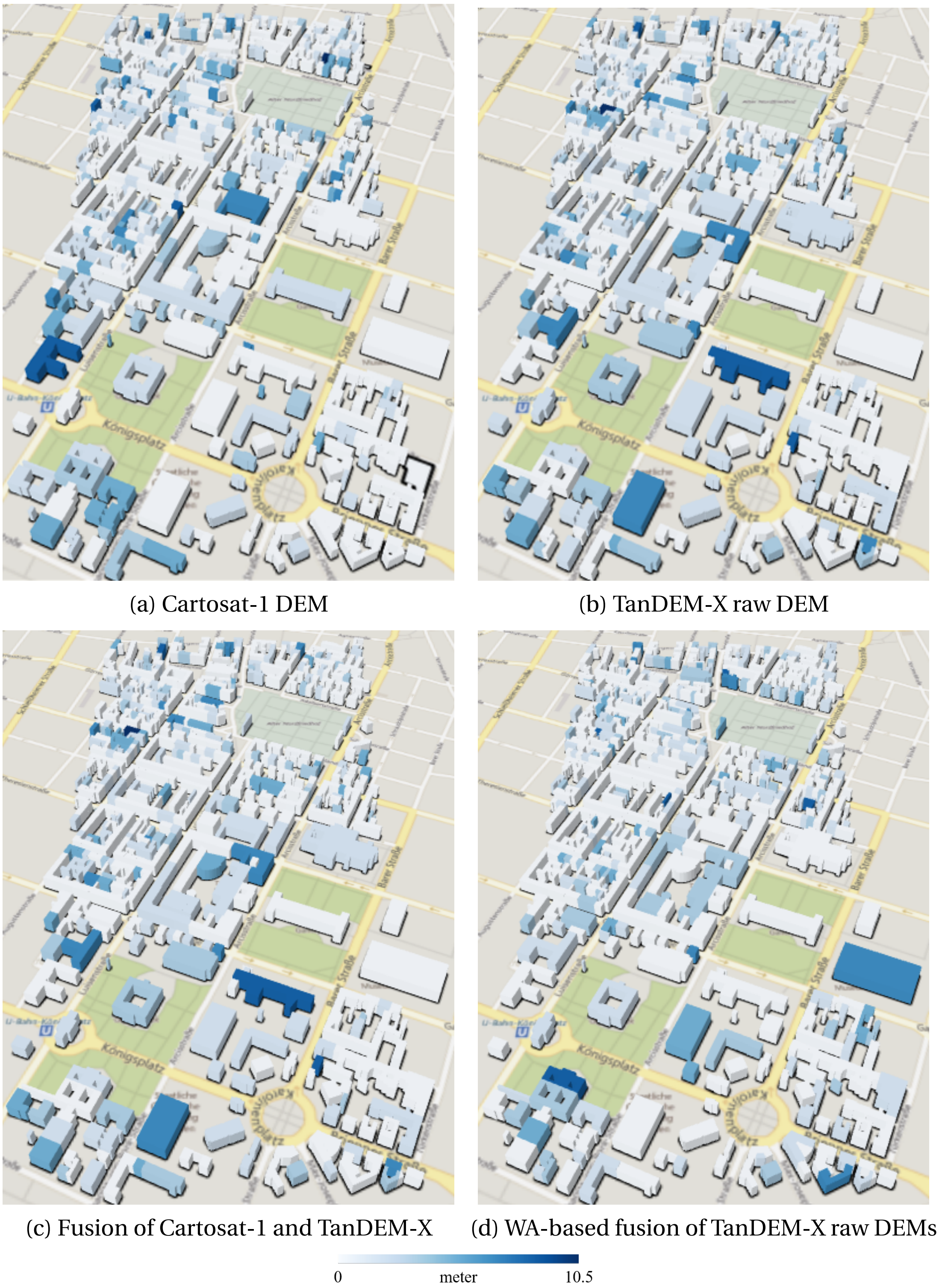
Table 3.14: Quantitative evaluations (in meter) of the LOD1 reconstructions of the urban scene using heights derived from different sources along with original building outlines of OSM

Input Heights	Median	RMSE	STD
Cartosat-1 (primary aligned DEM)	8.63	10.01	4.67
TanDEM-X raw DEM	9.68	10.16	4.28
ANN-based fused DEM: Cartosat-1 and TanDEM-X	9.56	9.97	4.28
WA based fused DEM: TanDEM-X	7.91	9.50	4.81
TV-L_1 based fused DEM: TanDEM-X	8.94	8.95	3.82
Huber based fused DEM: TanDEM-X	8.97	9.00	3.83
SAR-optical stereogrammetry: TerraSAR-X/WorldView-2	6.51	9.73	5.83

Table 3.15: Quantitative evaluations (in meter) of the LOD1 reconstructions of the urban scene using heights derived from different sources along with modified building outlines of OSM

Input Heights	Median	RMSE	STD
Cartosat-1 (primary aligned DEM)	-0.96	2.85	2.27
TanDEM-X raw DEM	-0.93	3.43	2.83
ANN-based fused DEM: Cartosat-1 and TanDEM-X	-0.92	3.09	2.48
WA based fused DEM: TanDEM-X	-0.72	2.81	2.5
TV-L_1 based fused DEM: TanDEM-X	-0.68	2.86	2.56
Huber based fused DEM: TanDEM-X	-0.67	2.96	2.64
SAR-optical stereogrammetry: TerraSAR-X/WorldView-2	-0.29	3.61	3.57

is needed to compute the height of the bottom and the top of a building along with the underlying terrain. Due to the limited the resolution of the height data utilized in this study, the focus did not lie on full 3D city model reconstruction but on simple prismatic building model reconstruction. For that purpose, the assumption of flat terrain at a constant height, which is valid in the selected study area was worked. For a complete 3D city model, more accurate measurements of the terrain and the bottom of building elevations would be necessary.



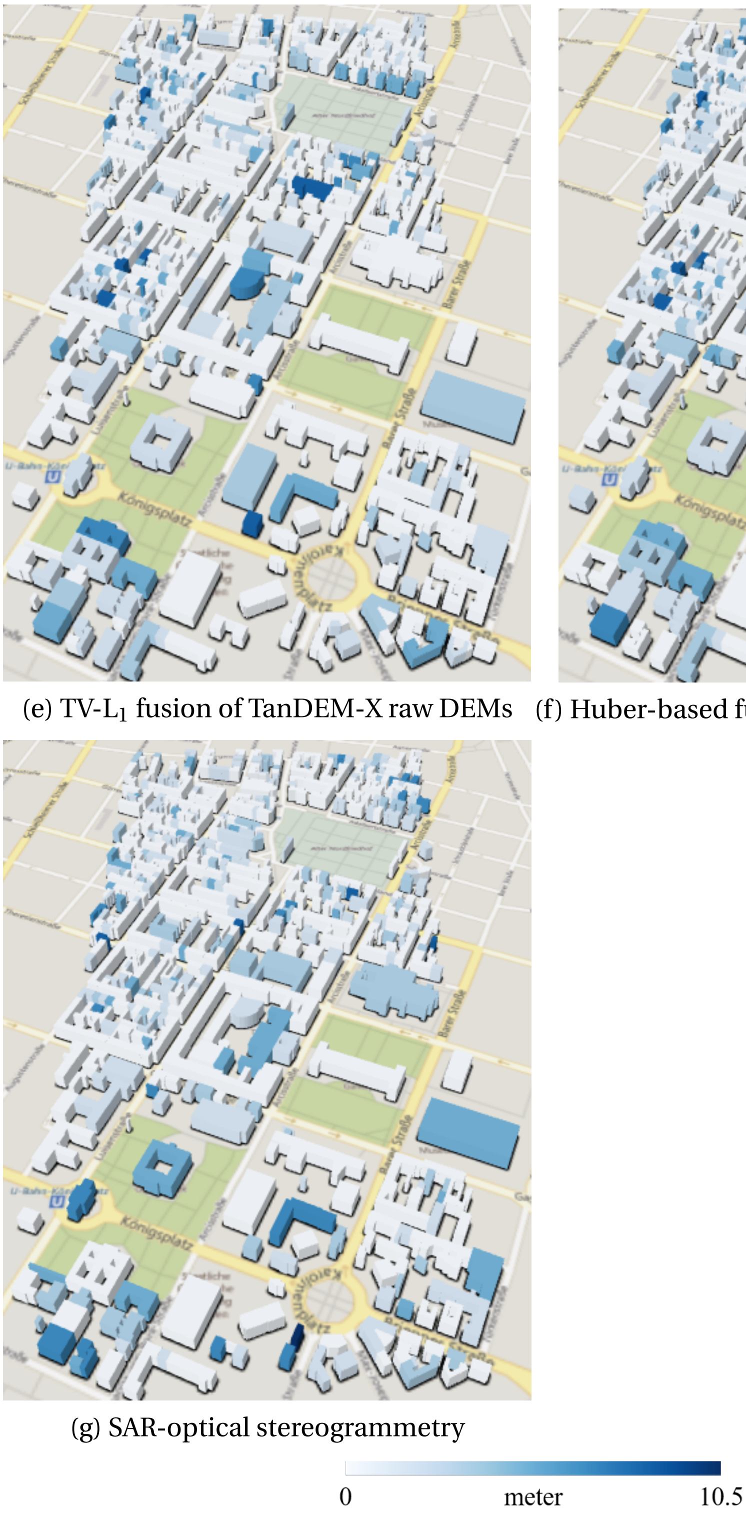
(a) Cartosat-1 DEM

(b) TanDEM-X raw DEM

(c) Fusion of Cartosat-1 and TanDEM-X

(d) WA-based fusion of TanDEM-X raw DEMs

Figure 3.22: LOD1 reconstructions of the study urban scene using heights derived from different sources and building outlines obtained from building foot prints layer of OSM. Colors indicate absolute height residuals



(e) TV-L_1 fusion of TanDEM-X raw DEMs (f) Huber-based fusion of TanDEM-X raw DEMs

(g) SAR-optical stereogrammetry

Figure 3.22: Continued

4 Conclusions and Outlooks

This dissertation summarized the results of investigations carried out according to the objectives delineated in Chapter 1. The main focus of these studies was to evaluate the potential of 3D reconstruction of urban areas using multi-sensor-derived data. To this end, data fusion techniques were employed to produce high-quality geospatial information such as elevations to ultimately reconstruct urban areas, especially buildings. In this regard, the following conclusions can be drawn:

- In this study, as a main objective, the potential of LOD1 3D reconstruction on a large scale was investigated using data exploited from remote-sensing-derived geodata and volunteered geographic information. For this purpose, heights provided for global urban mappings were applied such as heights generated through DEM fusion and SAR-optical stereogrammetry. Since building outlines as an essential requirement for 3D reconstruction cannot be accurately recognized in medium-resolution height sources, outlines provided by OSM were employed. It was also shown that the primary outlines were not perfect and should be modified and updated for accurate reconstruction. Finally, the median of heights within a building outline was calculated for LOD1 reconstruction. The final results demonstrated the possibility of prismatic building model generation (at the LOD1 level) on a large scale from easily accessible, remote sensing-derived geodata. The results also showed that the best LOD1 model was achieved using heights derived from the outputs of multi-sensor/multi-modal DEM fusion. The quantitative analysis also illustrated that the accuracy of LOD1 building modeling using heights provided by DEM fusion was better than 2.5 m i.e. less than the height of a building storey on average. In addition to DEM fusion, heights reconstructed from SAR-optical image pairs through stereogrammetry could potentially be applied for large scale LOD1 building modeling. However, the accuracy of the generated LOD1 model using heights achieved by SAR-optical stereogrammetry was lower than that of the LOD1 model produced using heights obtained by DEM fusion.

  As an outlook, the accuracy of the generated LOD1 models can be improved using other sources of information such as metadata (e.g., building heights, number of stories, etc.) provided along with VGI. In addition, the building heights for LOD1 modeling can be estimated using volunteered photos of buildings, i.e. single perspective images of buildings freely available on the Internet

- One of the potential remote sensing sources for 3D building reconstructions is elevations derived from medium-resolution DEMs such TanDEM-X. TanDEM-X DEM as a global DEM is not as perfect as optical-derived DEMs such as Cartosat-1 DEM. One possibility is to take advantage of data fusion techniques to fuse medium-resolution

DEMs. The output is a DEM with higher quality than primary input DEMs in urban areas. For this task, a typical data fusion technique, namely weighted averaging, is employed to pixel-wise fuse the elevation data. To achieve optimal fusion results, an innovative framework was developed to predict weight maps using a fully connected artificial neural network. The results demonstrated that the proposed method could efficiently improve the height precision of both Cartosat-1 and TanDEM-X DEMs up to 50% in urban areas and 22% in non-urban areas. Moreover, the proposed method was shown to increase the absolute accuracy through the data alignment. While the height precision improvement by the HEM-based method did not exceed 20% and 10% in urban and non-urban areas, respectively, if even HEMs would be available for DEM fusion.

Future research should consider the potential of using deep neural networks for DEM fusion. For example, a convolutional neural network can be designed for fusing DEMs, in which primary DEMs are considered as inputs and highly accurate elevations such as those provided by DEMs produced from high resolution LiDAR data are used as ground truth for training. The main advantage of this structure is to discard feature extraction and feature engineering steps for fusion. However, appropriately training a deep network demands quite a lot of training data.

- Another solution for improving the quality of medium-resolution DEMs such as TanDEM-X in urban areas is to use more sophisticated but efficient multi-modal DEM fusion techniques in the multiple TanDEM-X raw DEM mosaicking phase instead of simple WA. In this regard, it was proposed to apply TV-based variational models (TV-L_1 and Huber models) for TanDEM-X raw DEM fusion with the main focus on enhancing final DEM in urban areas, where building footprints are influenced by noise effects due to SAR imaging properties. For this purpose, different study subsets were selected from different land types which were explored over urban areas and surroundings. Apart from this, DEM fusion was investigated for raw DEMs with different geometries. At first, two nominal acquisitions with similar baseline configurations and HoAs were fused over different land types. In the next experiment, two raw DEMs with different HoAs were fused over a problematic terrain that suffering from PU errors. In the end, two DEMs were used with ascending and descending orbit directions along with different HoAs. In all the experiments, it was demonstrated that using variational models led to DEMs with high quality. A great performance of the Huber model was recorded for fusing two raw DEMs with different HoAs over the selected problematic area. Moreover, in urban areas, the variational models absolutely performed better than WA by reducing the noise effect and enhancing the outlines of buildings. However, the Huber model tended to provide a smoother fused DEM than TV-L_1. The results also demonstrated that the variational models, particularly TV-L_1, could significantly improve the quality of DEMs in comparison to WA. In addition, using the variational models was observed to improve the DEM quality by up to 2 m, particularly in inner-city subsets. In conclusion, carrying out TanDEM-X raw DEM fusion using the variational models with the ability to enhance building footprints and other useful high-frequency contents along with the ability to smooth the noise contributed to the production of a DEM with high quality.

- One other potential remote sensing source for generating elevations in urban areas is to use archive stored VHR SAR and optical imagery. This can be realized by developing a full 3D reconstruction framework based on the classic photogrammetric workflow. Accordingly, first, all prerequisites were analyzed for this task. The main requirement for SAR-optical stereogrammetry was to establish an epipolarity constraint to reduce the search space of the matching process. It was mathematically proved that the epipolarity constraint could be established for SAR-optical image pairs. Furthermore, the experimental analysis demonstrated that the epipolarity constraint could be employed for SAR-optical image pairs such as those from TerraSAR-X/WorldView-2. The analysis also showed that the epipolar curves were sufficiently straight. Because of the limited accuracy of RPCs delivered with optical data, the relative orientation between both SAR and optical images could be improved with respect to highly accurate SAR orientation parameters using multi-sensor block adjustment. This shifted the epipolar curves toward their correct positions. Then, the SGM-based dense matching algorithm was implemented using the MI similarity measure. The outputs were sparse point clouds with a median accuracy of about 1.5 to 2 m and the 25%-quantile of best points was in the sub-pixel accuracy domain. Overall, it is concluded that a 3D reconstruction framework can be designed and implemented for SAR-optical image pairs over urban areas.

  Future investigations are necessary to specifically design a similarity measure suitable for SAR and optical imagery, which is used as a cost function in the heart of SGM during dense matching. As a proposal, deep learning techniques can be applied for this purpose.

A Appendices

A.1 H. Bagheri, M. Schmitt, and X. X. Zhu. Fusion of TanDEM-X and Cartosat-1 elevation data supported by neural network-predicted weight maps. In: ISPRS Journal of Photogrammetry and Remote Sensing 144 (2018), pp. 285–297.

ISPRS Journal of Photogrammetry and Remote Sensing 144 (2018) 285–297

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Fusion of TanDEM-X and Cartosat-1 elevation data supported by neural network-predicted weight maps

Hossein Bagheri[a], Michael Schmitt[a], Xiao Xiang Zhu[a,b,*]

[a] Signal Processing in Earth Observation, Technical University of Munich, Munich, Germany
[b] Remote Sensing Technology Institute, German Aerospace Center, Oberpfaffenhofen, Wessling, Germany

ARTICLE INFO

Keywords:
TanDEM-X
Cartosat-1
Data fusion
Artificial neural network (ANN)
Digital elevation model (DEM)
ASTER GDEM
SRTM

ABSTRACT

Recently, the bistatic SAR interferometry mission TanDEM-X provided a global terrain map with unprecedented accuracy. However, visual inspection and empirical assessment of TanDEM-X elevation data against high-resolution ground truth illustrates that the quality of the DEM decreases in urban areas because of SAR-inherent imaging properties. One possible solution for an enhancement of the TanDEM-X DEM quality is to fuse it with other elevation data derived from high-resolution optical stereoscopic imagery, such as that provided by the Cartosat-1 mission. This is usually done by Weighted Averaging (WA) of previously aligned DEM cells. The main contribution of this paper is to develop a method to efficiently predict weight maps in order to achieve optimized fusion results. The prediction is modeled using a fully connected Artificial Neural Network (ANN). The idea of this ANN is to extract suitable features from DEMs that relate to height residuals in training areas and then to automatically learn the pattern of the relationship between height errors and features. The results show the DEM fusion based on the ANN-predicted weights improves the qualities of the study DEMs. Apart from increasing the absolute accuracy of Cartosat-1 DEM by DEM fusion, the relative accuracy (respective to reference LiDAR data) of DEMs is improved by up to 50% in urban areas and 22% in non-urban areas while the improvement by the HEM-based method does not exceed 20% and 10% in urban and non-urban areas respectively.

1. Introduction

Digital Elevation Models (DEMs) in diverse resolutions, levels of height accuracy and coverages are routinely produced by different techniques for a varied range of applications in different fields, such as navigation, geographical studies of the environment, or the ortho-rectification of remote sensing imagery. Particular attention is paid to the production of global DEMs, which represent homogeneous topography information for nearly all landmasses of the world. Different technologies have been employed for producing nearly global DEMs like the SRTM DEM (Rabus et al., 2003; Rodriguez et al., 2006), the ASTER GDEM (Tachikawa et al., 2011) or AW3D30 (Takaku et al., 2014; Tadono et al., 2014) which methodologically lie in two categories: SAR-interferometric and optical stereoscopic procedures. Each one of them has its own advantages and drawbacks that lead to DEMs with specific properties and limitations regarding final resolution and coverage. As an example, the SRTM DEM with a grid spacing of 1″ only covers the latitudes between 56°S and 60°N. An example for an elevation model derived from optical stereo data is the AW3D30 DEM based on ALOS PRISM data, which provides both higher accuracy and larger coverage (between 82°S and 83°N) than the SRTM DEM, but contains some void areas due to missing information caused by clouds, snow, etc. (Takaku et al., 2016).

Recently, a new global topography dataset was attained through the TanDEM-X mission, which provides a spatial resolution of 12 m with coverage of nearly the whole earth. The TanDEM-X mission comprises twin SAR satellites (TerraSAR-X and TanDEM-X launched in June 2007 and June 2010, respectively), which fly in adjacent orbits to acquire bistatic SAR images. The mission was devised to produce DEMs with a target accuracy according to High-Resolution Terrain Information standard level 3 (HRTI-3) Heady et al. (2009): i.e., with a relative height accuracy finer than 2 m for areas including slopes lower than 20%, and 4 m for slopes steeper than 20% (Krieger et al., 2007). The special satellite constellation equipped with X-band SAR sensors exploits a bistatic SAR interferometry configuration with single pass acquisitions free of atmospheric and temporal decorrelation effects and consequently provides the first high-resolution global DEM. The initial DEM product, the so-called raw TanDEM-X DEM with nominal pixel

[*] Corresponding author at: Remote Sensing Technology Institute, German Aerospace Center, Oberpfaffenhofen, Wessling, Germany and Signal Processing in Earth Observation, Technical University of Munich, Munich, Germany.
E-mail address: xiao.zhu@dlr.de (X.X. Zhu).

https://doi.org/10.1016/j.isprsjprs.2018.07.007
Received 14 October 2017; Received in revised form 2 April 2018; Accepted 17 July 2018
0924-2716/ © 2018 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

spacing of 0.2 arcsec (6 m at the equator), is the output of the Integrated TanDEM-X Processor (ITP) (Fritz et al., 2011). The raw TanDEM-X DEM is finally cast in a grid with pixel spacing of 0.4 arcsec (12 m at the equator) after DEM calibration (Gruber et al., 2012; Rossi et al., 2012) and mosaicking (Gruber et al., 2016) to obtain the global DEM according to HRTI-3 standard. While the standard DEM globally represents non-urban areas with unprecedented relative accuracy (Zink et al., 2014), the drop of the DEM's spatial resolution makes the final standard DEM unsuitable for high-resolution 3D reconstruction in urban areas (Rossi et al., 2013a). Consequently, the raw TanDEM-X DEM provides a more spatially detailed mapping of urban areas in comparison to the standard version of the global TanDEM-X DEM. However, preliminary visual inspection of raw TanDEM-X DEM data still indicates unfavorable spatial resolution and drop of height precision, especially for areas with topographically difficult surfaces—like urban areas (Rossi et al., 2011)—and the requirement for TanDEM-X quality enhancement in these areas.

One solution for refining the TanDEM-X DEM in difficult terrains can be a fusion with elevation data derived from other sources with different acquisition properties. Optical imagery, e.g., does not suffer from SAR-intrinsic imaging effects, such as layover and shadowing, which influence the appearance of InSAR DEM products.

Optical DEMs result from stereoscopic 3D reconstruction of high-resolution optical images. For example, Cartosat-1 data provides a series of DEMs with relative accuracy of HRTI-3 standard (2–3 m). Cartosat-1 (also called IRS-P5) is an Indian satellite (launched in May 2005) equipped with a pushbroom sensor consisting of an ensemble of CCDs with a size of 2.5 m in two lines for along track scanning of scenes with a stereo angle of 31° (Srivastava et al., 2007). It is particularly intended to produce a high-resolution DEM with coverage of a relatively wide area (Ahmed et al., 2007), and is used, for instance, for large-scale DEM generation in Europe (Uttenthaler et al., 2013). The Cartosat-1 data are provided with Rational Polynomial Coefficients (RPCs) computed from the mission's orbit and attitude information. Evaluations have demonstrated that their accuracy – for instance measured by Root Mean Square Errors (RMSE) – is restricted to multiple hundred meters (Lehner et al., 2007) i.e. the final produced DEM in spite of fairly high relative accuracy is absolutely located in an incorrect position. The poor accuracy of the RPCs affects the stereo intersection results and causes residuals in the final DEM product. Generally, a good distribution of Ground Control Points (GCPs) is needed for RPC refinement and bias compensation (Teo, 2011) of high-resolution optical images like those provided by Cartosat-1, but availability of GCPs cannot always be ensured. The conventional solution is to use available global DEMs—like the SRTM DEM—as an external vertical reference for bias compensation and RPC refinement (Kim and Jeong, 2011). The emergence of the TanDEM-X DEM as global DEM of HRTI-3 standard, as opposed to the SRTM DEM of DTED-2 standard (Krieger et al., 2007), provides the required height reference with higher accuracy for refining the RPCs of Cartosat-1 that will ultimately result in more accurate Cartosat-1 DEM.

Considering the aforementioned defects of the Cartosat-1 and TanDEM-X elevation data, the main objective of this paper is to develop a framework for efficient fusion of TanDEM-X and Cartosat-1 DEM over urban areas. Eventually, this fusion will increase the height precision of the final DEM over urban areas, while its absolute vertical accuracy is improved to the level of the TanDEM-X DEM.

Data fusion approaches with great deal of applications in remote sensing can be adapted for DEM fusion tasks (Schmitt and Zhu, 2016), and for this aim, various methods have been investigated for different kinds of DEMs. Reinartz et al. (2005) employed weighted averaging for the fusion of SPOT-5 and SRTM DEMs. In a similar study (Roth et al., 2002), weighted averaging was used to fuse ERS TanDEM data and SRTM data with MOMS-2P data. A more advanced technique was proposed by Papasaika et al. (2011), in which sparse representation supported by weights served for fusion of DEMs from various data sources. Pock et al. (2011) proposed Total Generalized Variational (TGV) methods for fusion of airborne optical-stereoscopic DEMs, while a weighted version of total variational (TV) method and TGV were examined by Kuschk et al. (2017) on different space borne optical DEMs. Fuss et al. utilized the modified K-means clustering algorithm to fuse multiple overlapping radargrammetric Envisat-2 DEMs (Fuss et al., 2016). The first experience for fusion of TanDEM-X and Cartosat-1 DEMs over urban and non-urban areas comes with Rossi et al. (2013b), in which only the TanDEM-X DEM was improved over non-urban areas by weighted averaging and prior knowledge of DEM qualities.

Among the aforementioned methods, weighted averaging (WA) is wellknown and frequently used for DEM fusion purposes (Gruber et al., 2016; Reinartz et al., 2005; Roth et al., 2002; Rossi et al., 2013b; Bagheri et al., 2017a; Deo et al., 2015), because of its simple implementation and low computational cost. In addition, most advanced techniques apply weights to assist the fusion process to reach the desired output. This means the weights play a key role for efficient fusion of DEMs, especially in the case of multi-sensor DEM fusion, like stereoscopic-optical and InSAR DEMs (Bagheri et al., 2017a). The key problem with using DEM fusion approaches, especially for WA, is applying appropriate weight maps receptive to each DEM—which used to be proportional to the expected height residuals. For this purpose, prior knowledge about existing DEM errors will always be beneficial for the fusion process. One solution for predicting the expected errors is based on an error propagation analysis through the DEM generation procedure. However, usually such a model can only be an approximation and may not model all potential error sources.

An alternative is to learn the error patterns by comparing exemplary areas of interest and corresponding ground truth reference data: e.g., derived from high-precision LiDAR measurements. This way, suitable weights can be predicted for newly incoming datasets for which neither detailed information about the height errors nor any ground truth data are available.

This paper is an extension of Bagheri et al. (2017a,b), in which we mostly focused on evaluating the accuracy of Cartosat-1 and TanDEM-X DEMs with respect to high-resolution LiDAR reference data to provide a judgment about the potential of the Cartosat-1 and TanDEM-X DEM fusion over urban areas. The evaluation confirmed that the TanDEM-X DEM quality over urban areas is not ideal in comparison to the Cartosat-1 elevation data, and that data fusion can improve the quality of the final DEM product.

In this paper, a sophisticated framework for appropriate weight map prediction is proposed (Section 4.1). Firstly, suitable spatial features, along with height residuals, are extracted from the training DEMs. After that, the pattern of refined height errors in relation to features are learned by an artificial neural network (ANN) to predict weights for WA DEM fusion. In order to provide a baseline result, first simple DEM fusion using the weights derived from the TanDEM-X Height Error Map (HEM) and the Cartosat-1 matching standard deviations are shown and discussed in Sections 4 and 5. The final results (Section 5) illustrate the ANN-supported DEM fusion can improve the quality of both DEMs over urban areas to generate a global high-resolution DEM in contrast to standard HEM-based weights. Finally, fusion framework is validated using SRTM DEM and the ASTER GDEM data to investigate the possibility of transferring the proposed approach to DEMs of other specifications as well (Section 7).

2. Study area and DEMs

The data used for the experiments described in this paper were acquired over the area of Munich, Germany. The Cartosat-1 DEM with nominal grid size of 5 m was generated from stacks of overlapping images by the dense matching and 3D stereoscopic reconstruction toolboxes embedded in the XDibias image processing system of DLR. The description of the DEM generation procedure has been detailed in d'Angelo et al. (2008). The TanDEM-X raw DEM produced by DLR's

Table 1
Properties of raw TanDEM-X tile.

Raw TanDEM-XDEM	
Center incidence angle	38.25°
Equator crossing direction	Ascending
Look direction	Right
Height of ambiguity	45.81 m
Total number of looks	22
Pixel spacing	0.2 arcsec
HEM mean	1.33 m

Table 2
Properties of Cartosat-1 tile. * For more information, look up (The Federal Agency for Cartography and Geodesy of Germany (BKG), 2016).

Cartosat-1 DEM	
Stereoscopic angle	31°
Max number of rays	11
Min number of rays	2
Horizontal reference	BKG orthophotos*
Vertical reference	SRTM DEM
Pixel spacing	5 m
Mean height error (1σ)	2–3 m

Integrated TanDEM-X Processor (ITP), in which bistatic SAR data-takes acquired in strip map mode are processed interferometrically. It is delivered with a grid spacing of 0.2 arcsec. More details about the used TanDEM-X and Cartosat-1 tiles are collected in Tables 1 and 2.

For representative experiments, we chose study subsets representing 6 different land types that are usually found over urban areas and their surroundings (see Fig. 1). The main characteristics of these land types are briefly stated in Table 3. The naming abbreviations are meant to facilitate referencing each subset throughout this paper.

Both for training and evaluation of the ANN-fusion framework, highly accurate reference elevation models are provided by high-resolution LiDAR point clouds with a density of one point per square meter.

3. Data preparation and alignment

Before implementing the fusion process, the Cartosat-1 and TanDEM-X data should be homogenized in terms of horizontal and vertical references, as well as pixel spacing. The initial coordinate systems and the pixel spacing of the study DEMs are expressed in Table 4. The final DEM specifications identify the reference systems and nominal pixel spacing of all used DEMs after the data preparation process. To prepare the data for the actual fusion, the Cartosat-1 and the TanDEM-X DEMs are resampled to 5 m pixel spacing, which is nearly identical to their own initial resolutions and grid spacing.

In addition, after preparing the elevation data in the form of the desired grid and reference, the tiles of study DEMs (TanDEM-X and Cartosat-1) should be aligned together to compensate any rotational and translational discrepancies. Usually, the best operational way for this purpose would be to utilize the well calibrated TanDEM-X data as an external DEM for a refinement of the Cartosat-1 RPCs. However, since this study does not focus on Cartosat-1 DEM generation and only starts with an available Cartosat-1 DEM product, DEM coregistration needs to be carried out. For this purpose, the ICP (Iterative Closest Point) algorithm is used to align the Cartosat-1 DEM tile to the TanDEM-X DEM tile (Ravanbakhsh and Fraser, 2013).

4. TanDEM-X and Cartosat-1 DEM fusion

The results of relative and absolute accuracy assessment presented in Bagheri et al. (2017a) demonstrated that over urban areas the height precision and resolution of the TanDEM-X DEM is not as fine as that of the Cartosat-1 DEM, whereas both DEMs have nearly identical quality over non-urban areas such as agricultural and forested fields. On the other hand, the TanDEM-X is absolutely more accurate than Cartosat-1 DEM. As explained in Section 1, data fusion techniques can be applied to take advantage of the properties of both kinds of DEMs to finally reach a DEM product with higher precision and absolute accuracy.

While the simplest method for DEM fusion is weighted averaging, its main challenge is to employ suitable weight maps respective to each DEM. For example, the TanDEM-X and Cartosat-1 data can be fused by using weight maps delivered from the provided height error maps using simple weighted averaging:

$$\mathbf{D}_F = \mathbf{W}_T^n \odot \mathbf{D}_T + \mathbf{W}_C^n \odot \mathbf{D}_C \tag{1}$$

where $\mathbf{W}_T$ and $\mathbf{W}_C$ are the normalized weights of the TanDEM-X and the Cartosat-1 heights, respectively, $\mathbf{D}_T$ and $\mathbf{D}_C$ are the height values taken from the TanDEM-X and Cartosat-1 DEMs, $\odot$ denotes the element-wise product and $\mathbf{D}_F$ refers to final fused DEM.

Usually, attached to InSAR DEMs such as TanDEM-X data, an additional product (called Height Error MAP: HEM) is provided, which roughly describes the quality of the generated DEM by the interferometric process based on coherence analysis (Martone et al., 2012), the number of the looks and the baseline configuration (Just and Bamler, 1994). Similarly, For optical DEMs such as Cartosat-1 DEM, the quality map (also can be called HEM) is computed from stereo matching analysis. Both HEMs are produced by error propagation analysis through the chain of DEM generation from the source data takes.

The main disadvantage of HEM-based fusion is that HEMs are not always available in the form of DEM metadata, which holds in particular for DEMs generated from optical stereo data. This limits a broader applicability of HEM-based DEM fusion.

For DEM fusion, pixel-wise weight maps should be created from the HEM maps. Two strategies can be pursued for weight map generation. The first one is to calculate weights as the inverse proportional of the squared height errors e_i:

$$w_i = \frac{1}{e_i^2} \tag{2}$$

Another way is to compute weights from the normalized residuals e_{in}:

$$w_i = 1 - e_{in} \tag{3}$$

4.1. DEM fusion support by neural network-predicted weights

As an alternative to the standard weight map generation process, in this paper we propose applying supervised learning that exploits high-resolution LiDAR ground truth data available for selected areas as training data to predict the height residual patterns and the corresponding weights for test data subsets. Fig. 2 displays the framework of the proposed DEM fusion algorithm. In the heart of the proposed framework, an ANN is used to learn the relationship patterns of height errors and corresponding DEM features, which can subsequently be used for forecasting weight maps. The proposed framework (Fig. 2) can be summarized in three main steps:

1. spatial feature extraction from DEMs and height error calculation
2. data refinement
3. (a) training the ANN on dedicated training subsets for which ground truth data is available and (b) applying the ANN parameters to target subsets.

The output of the ANN is a predictive model that works as a weight predictor in target areas to which DEMs are fused based on the patterns explored in training subsets. More details of the framework's steps will be explained in the following subsections.

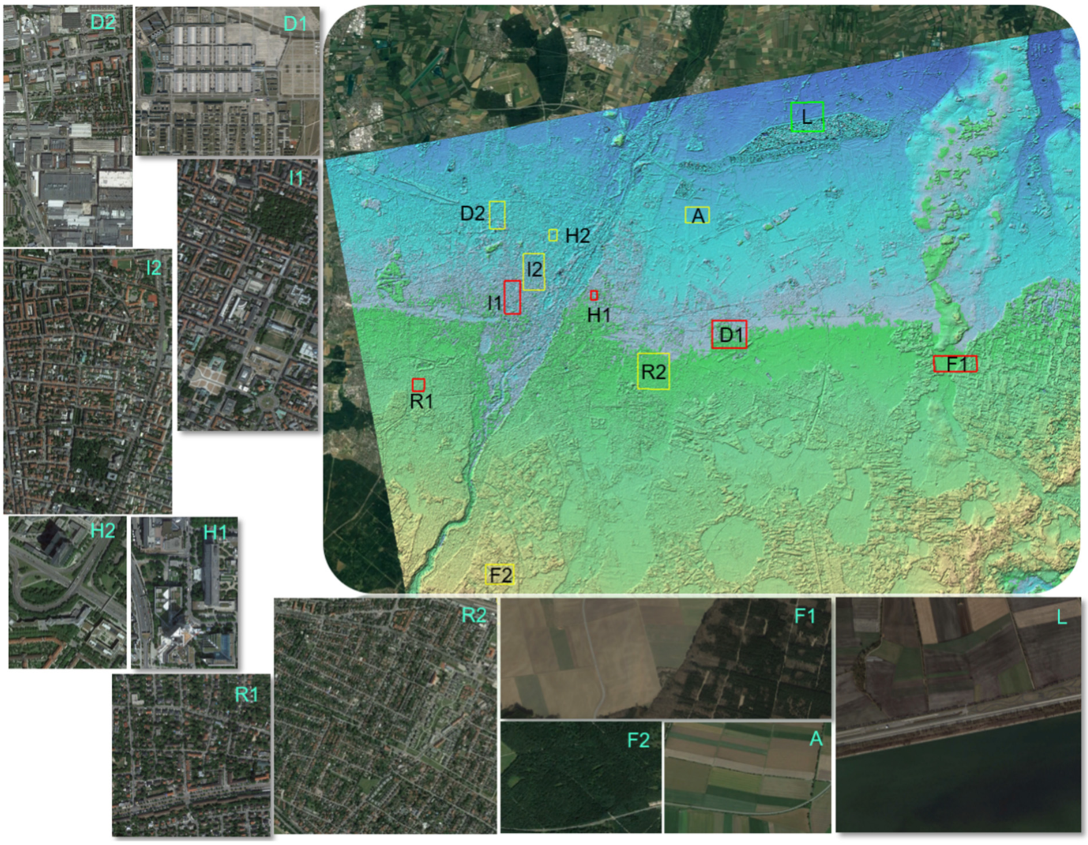


Fig. 1. Display of land types and location of study subsets from Munich area.

Table 3
The main properties of study subsets.

Subsets	Descriptions of study areas Characteristics	Naming	Areas (km^2)
Industrial	Open mid rise rectangular buildings	D1, D2	1.77, 0.78
Inner city	Compact mid rise with complicated shape and perhaps relatively high	I1, I2	0.92, 1.384
High building	High rise buildings and skyscrapers	H1, H2	0.09, 0.15
Residential	Open low rise buildings for single family	R1, R2	0.26, 2.12
Forested	Dense canopy of trees	F1, F2	0.54, 1.0
Agricultural	Classical land farms	F1, A	0.68, 0.66

Table 4
Specifications of initial DEMs and final DEM, output of preparation step.

	Horizontal reference	Vertical reference	Pixel spacing
Cartosat-1	UTM(WGS84)	EGM96	5 m
TanDEM-X	WGS84	WGS84	0.2 arcsec(~6 m)
LiDAR	Gauss Krüger(Bessel)	Bessel	1 m
Final DEM	UTM(WGS84)	WGS84	5 m

4.1.1. Feature extraction and height residual computation

For the training of the ANN, training data are selected from representatives of different land types that can usually be observed over urban areas. The description of these land types are in Section 3. From those, different kinds of spatial features describing landscaping and roughness properties of the land surface are extracted. Several studies clarify the relationship between the spatial features and DEM qualities (Toutin, 2002; Papasaika et al., 2009; Reinartz et al., 2010). The following spatial features are useful for DEM fusion (Olaya, 2009):

- Geometrical parameters such as (1) Slope, which expresses the maximal rate of varying heights and (2) Aspect, which is the direction of the steepest slope in the mask window.
- Statistical land surface parameters like (1) Anisotropic Coefficient of Variation (ACV), which describes the general geometry of local surfaces for distinguishing elongated and oval landforms; (2) Topographic Ruggedness Index (TRI), which is the 2D standard deviation filter; (3) Topographic Position Index (TPI), which is the difference between height of a pixel and mean height of neighboring pixels; (4) Roughness, which is the largest height difference of target pixel and its surrounding cells; (5) Ruggedness, which is defined as the range value within an area; (6) Surface Roughness Factor (SRF), which is related to the normals to land surface; and (7) Entropy, which determines the uncertainty of height estimation in the selected window.

In addition to these parameters, edge values can also be extracted pixel-wisely by common edge detectors like the Sobel filter. At last, the HEM delivered with the TanDEM-X DEM and the quality map of Cartosat-1 DEM can also be used as a feature that reflects one source of induced errors in DEM. Apart from the HEM and the quality map, all mentioned features are extracted by convolution with a 3 × 3 square window as a mask around each cell. Fig. 3 exemplarily shows the maps for these features extracted from the Cartosat-1 data in the industrial area, subset I1.

Moreover, height residual maps are calculated for all training subsets by subtracting the LiDAR ground truth elevations from the

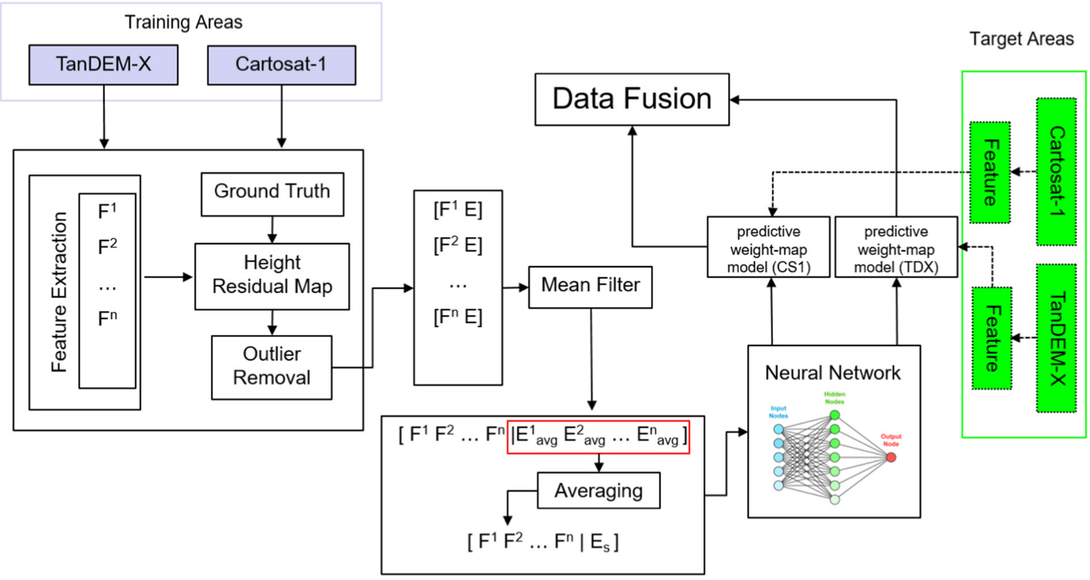


Fig. 2. The framework designed to estimate the adaptive weights by ANN for the TanDEM-X and Cartosat-1 DEM fusion.

corresponding DEM elevations. The obtained feature maps, along with these height residuals, are used to train an ANN to model the relationship between feature values and height errors. Fig. 5 exemplarily depicts the variation of errors with respect to feature values in the subset of the industrial area (I1) for both study DEMs. The height residuals, at first, are very noisy, so that a low pass filter (details explained in the following subsection) is required to reveal the error patterns as depicted in Fig. 5.

4.1.2. Data refinement

Prior to constituting the ANN structure, another important step is to refine the height errors oriented to extracted spatial feature values to get rid of outliers and decrease the noise effects. The calculated height residuals are polluted by high-frequency noise, which will affect the training of the ANN.

The performance of the network in the case of using smoothed residual maps derived from the refinement step, as well as using raw data without smoothing, only removing the outliers, are illustrated in Fig. 4. Fig. 4b indicates that noisy height residuals disrupt the training procedure and prevent the ANN from recognizing the error patterns. The correlation evaluation of desired outputs and results achieved by ANNs demonstrates the efficiency of the refinement step. By employing the refinement framework, the networks can learn the error pattern respective to features with high correlation (more than 0.98) between outputs of training and target values for TanDEM-X and Cartosat-1 DEMs, respectively. Without implementing the refinement, the training performances of networks are significantly lower, illustrated by an output with correlations lower than 0.50 and 0.35 for the TanDEM-X and the Cartosat-1 DEM, respectively.

To reduce the noise effects with the aim of promoting the training, a smoothing process, characterized by two-step mean filtering is carried out. The first step of the refinement is to bin the feature values that can be obtained by a simple empirical-statistical binning technique. The feature values f_i^j correspond to the height residuals e_i by:

$$\begin{bmatrix} f_1^1 & f_1^2 & \cdots & f_1^n \\ f_2^1 & f_2^2 & \cdots & f_2^n \\ \cdots & \cdots & \cdots & \cdots \\ f_m^1 & f_m^2 & \cdots & f_m^n \end{bmatrix} \Leftrightarrow \begin{bmatrix} e_1 \\ e_2 \\ \cdots \\ e_m \end{bmatrix} \tag{4}$$

where f_i^j is the value of the feature $j \in \{1, 2, \ldots, n\}$ and e_i is corresponding height residual value in pixel $i \in \{1, 2, \ldots, m\}$.

At first, errors exceeding 3 × NMAD (e_ζ, which are identified by index ζ) are detected as outliers and then eliminated along with their corresponding feature values $\left[f_\zeta^1 \;\; f_\zeta^2 \;\; \cdots \;\; f_\zeta^n \right]$ from the training dataset. The normalized median absolute deviation (NMAD) is recommended as a robust accuracy measure rather than the classical root mean square error (RMSE) for the mitigation of outliers affecting the elevation data of the study DEMs (Höhle and Höhle, 2009).

The relation (4) can be rewritten in the form of feature vectors that include the values of each feature type for all pixels of the DEM:

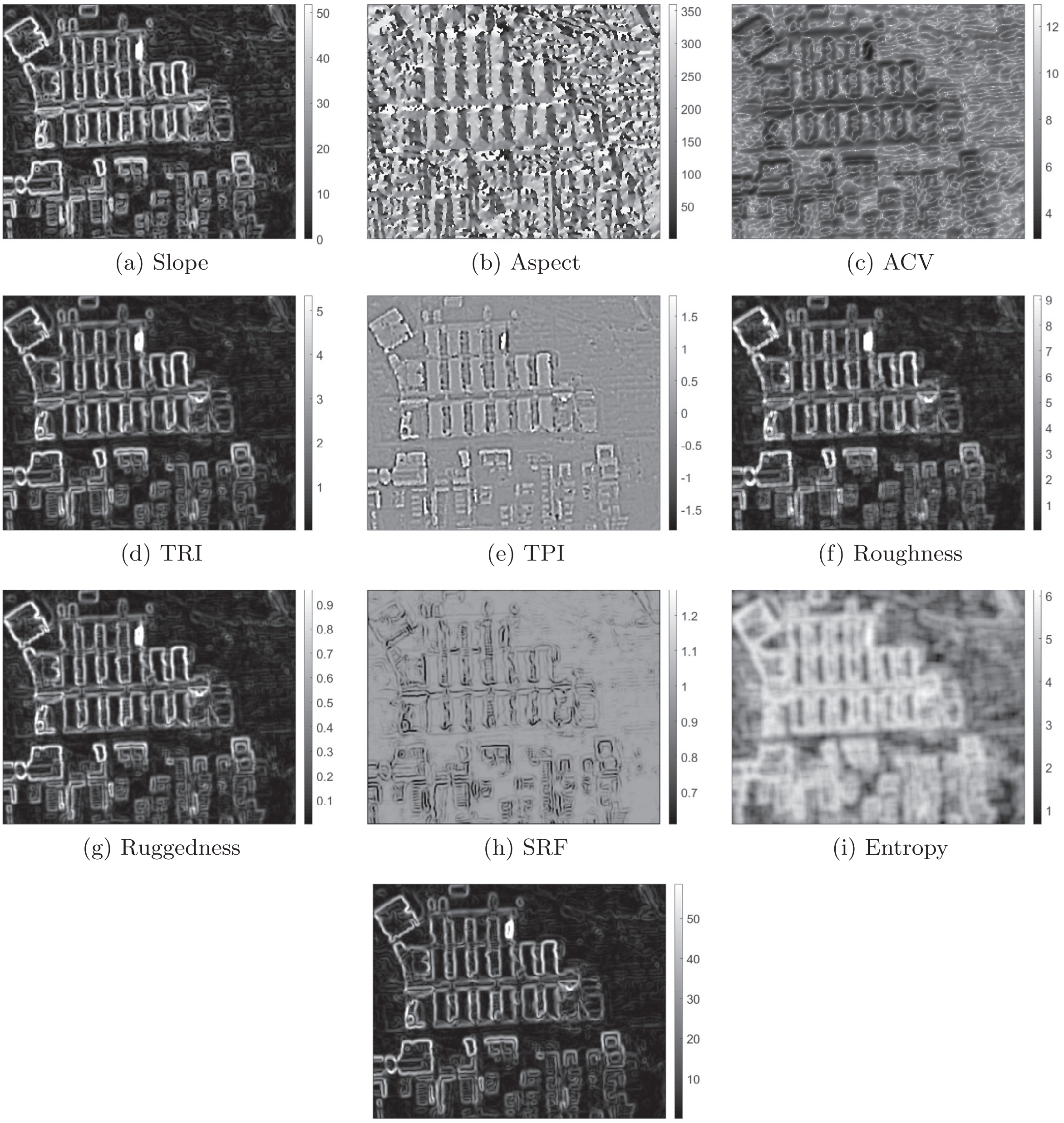
$$\begin{aligned} &[\mathbf{F^1} \;\; \mathbf{F^2} \;\; \ldots \;\; \mathbf{F^n}] \Leftrightarrow \mathbf{E} \\ &\mathbf{F^j} = [f_1^j \;\; f_2^j \;\; \cdots \;\; f_m^j]^T \end{aligned} \tag{5}$$

After removing outliers, The values of the feature vector ($\mathbf{F^j}$) and their corresponding height residuals $\mathbf{E}$ are binned by the Freedman-Diaconis rule (Birgé and Rozenholc, 2006):

$$N = \frac{f_{max}^j - f_{min}^j}{h} \tag{6}$$

where $h = 2 \times I \times k^{-1/3}$. The output of above formulation is the number of bins (N) for feature j with bin width of h, just by detecting the max and min values of measured feature (f_{max}^j and f_{min}^j). I is the interquartile range and k is the number of measurements that are remaining height residuals after outlier removal. In other words, k refers to number of pixels whose height errors (e_i) are lower than the threshold 3 × NMAD. The mean filter is applied bin-wise to generate smoother height residual. The output of this filtering is the numerical feature-error model in which each feature vector $\mathbf{F^j}$ corresponds to a new smoothed height residual map $\mathbf{E_{avg}^j} = [e_{1avg}^j \;\; e_{2avg}^j \;\; \cdots \;\; e_{mavg}^j]^T$. It has to be noted that infrequent feature values are thrown away by a threshold. This procedure should be followed for each type of feature. Fig. 5 presents the graphical depictions of feature-error models derived for an industrial area (subset I1) for TanDEM-X and Cartosat-1 after binning and mean filtering. Consequently, for each pixel, n height residual values at last are acquired. This means there are n residual maps, which are linked to n numerical feature-error models ($[\mathbf{F^1} \;\; \mathbf{F^2} \;\; \ldots \;\; \mathbf{F^n} \;|\; \mathbf{E_{avg}^1} \;\; \mathbf{E_{avg}^2} \;\; \ldots \;\; \mathbf{E_{avg}^n}]$).

Next, the second step of the smoothing process is to average again the achievements of the former step (smoothed height residuals) to



(a) Slope (b) Aspect (c) ACV

(d) TRI (e) TPI (f) Roughness

(g) Ruggedness (h) SRF (i) Entropy

(j) Edginess

Fig. 3. Feature maps extracted from DEM from industrial area (I1) of Cartosat-1 DEM.

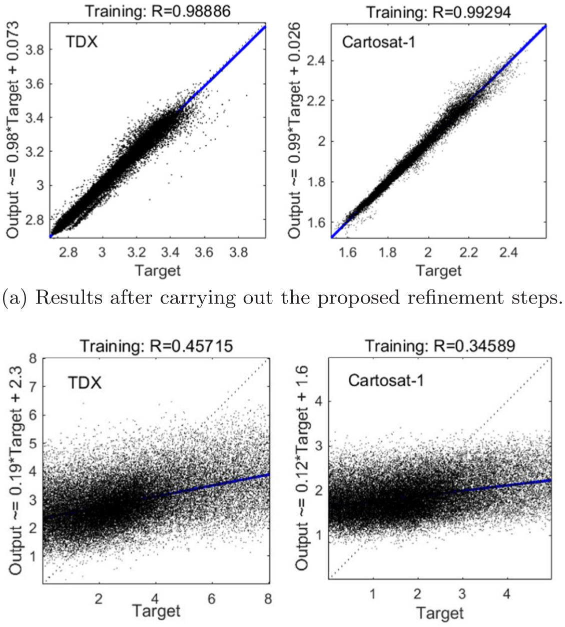
finally create a unique height residual. After this refinement, the data are ready to insert into the ANN for training and exploring the patterns.

4.1.3. Weight map generation by ANN

The filtered outcomes from the previous stage are employed to train a fully connected feed-forward neural network. After feature extraction, height residual computation and refinement based on the pipeline described in Section 4.1.2, the outputs become the input into the ANN. The ANN is trained using the filtered feature vectors as inputs and the modified height residuals as outputs, which are cast in the form of:

$$[\mathbf{\Phi_1} \;\; \mathbf{\Phi_2} \;\; \ldots \;\; \mathbf{\Phi_m} \;\mid\; \mathbf{E_s}],$$
$$\text{where} \quad \mathbf{\Phi_i} = [f_i^1 \;\; f_i^2 \;\; \ldots \;\; f_i^n]^T, \quad i \in \{1, 2, \ldots, m\} \tag{7}$$

contains the values of the different features for a given pixel i and E_s is the final smoothed residual map obtained through the two-step mean filtering. Fig. 7 shows the structure of the network, which consists of an input layer in which neurons with the label of the feature values of each pixel ($\mathbf{\Phi_i}$) are connected to the smoothed height residual of the corresponding pixel through the hidden layers. In the repetitive process, with back propagation training, the weights of neurons are gradually modified to decrease the discrepancy between smoothed height residual maps and the map achieved by the network. The main achievement of



(a) Results after carrying out the proposed refinement steps.

(b) Results without using the proposed refinement.

Fig. 4. Regression plot of training data for industrial area (Sub D1).

the ANN after successful training is that a model can estimate the weight maps for each part of DEMs from forecasting the height residuals just by measuring the spatial features.

The optimal structures of ANNs for both study DEMs were investigated by tracking the cost function values during the training stage. The performances of networks were evaluated by considering one hidden layer and changing the number of neurons in this layer. Then, deeper networks were examined by adding another hidden layer. Fig. 6a and b depicts the performance of the neural networks on test data for an increasing number of neurons in the ANNs with one hidden layer, as well as for structures designed with two hidden layers. The plots display that the rise in the number of neurons in the first layer is more influential than adding the layers and making the networks deeper. In ANNs with one hidden layer, the general trend of changing costs remained stable after adding more than 20 neurons. In other words, adding more neurons to a hidden layer does not change the performance of the networks and the subsequent DEM fusion result. This evaluation systematically determined the optimal structure of NNs considered for both DEMs.

5. TanDEM-X and Cartosat-1 DEM fusion results

One additional advantage of Cartosat-1 and TanDEM-X DEM fusion is to increase the absolute vertical accuracy with respect to the data accuracy provided by the original Cartosat-1 DEM. By successful alignment of the Catosat-1 and TanDEM-X DEMs, the Cartosat-1 DEM is vertically positioned in the location of TanDEM-X, which indicates an improvement of the absolute geolocation through DEM fusion. After vertical alignment, the HEM- and ANN-based approaches are examined to increase the height precision of both DEMs.

HEM-based fusion was implemented for all subsets except subsets F1 and L, due to unavailability of values of STD of matching for these areas. The HEM-based fusion results for the other subsets are presented in Table 5. The common metric for measuring the accuracy of DEMs is root mean square error (RMSE). Additionally, Normal Median Absolute Deviation (NMAD) is used as another measure for height precision analysis (Höhle and Höhle, 2009). The RMSE measure reflects the effects of whole larger and small errors but NMAD implies the presence of small errors in elevation data.

The results indicate that using the HEM of the TanDEM-X data and STD of matching for Cartosat-1 DEM could only increase the height precision of TanDEM-X elevations and an improvement for Cartosat-1 data cannot always be fulfilled for all land types.

In ANN-based fusion approach, training data are selected from diverse land types mentioned in Table 3, such as industrial, inner city, residential areas, high building subset, agricultural and forested areas. The usage of training data from different land types guarantees the presence of all possible values of features respective to height residuals in the process of pattern recognition by the NN, and give the assurance of discovering a more general model that can be used for any arbitrary land type. After successful training, the ANN can be applied for predicting the height residual in selected target areas where two DEMs are supposed to be fused. The predicted residual maps are used as weight maps in the weighted average fusion. Thus, two separate ANNs are needed for both the TanDEM-X and Cartosat-1 DEM to generate individual weight maps for each DEM separately. For the experiments in this paper, two strategies are followed regarding the combination of training subsets.

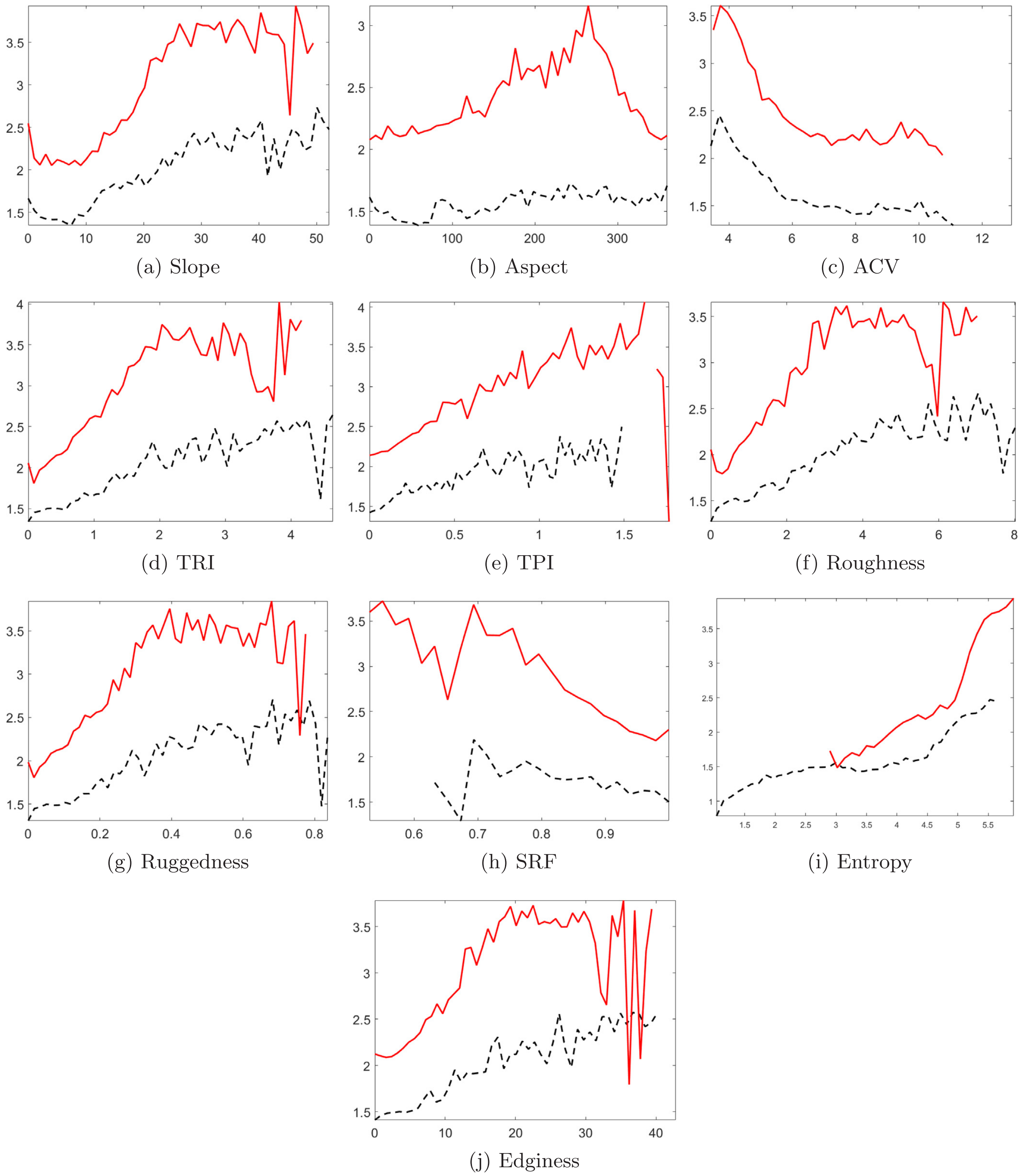
In strategy A, separate ANN are trained for each specific land type in order to provide class-specific weight map predictions. In this case, the subsets D1, I1, H1, R1 as well as F1 separated into its forested and its agricultural segments are used as training subsets, on which six different ANNs are trained. The resulting weight map predictors that are used for DEM fusion in the respective target areas are then applied on the corresponding test areas D2, I2, H2, R2, F2 and A.

In strategy B, all subsets of all different land types are simultaneously used as training data to create a general predictor model that can be used for weight map generation in arbitrary target subsets. In this experiment, the subsets D1, I1, H1, R1 and F1 are used to train the ANN, and the resulting output model is used for all target subsets.

In both experiments, 70% of data from the training subsets are devoted to training, and 15% are for validation to control the training process in order to avoid over-fitting and under-fitting. However, the whole process of the proposed framework will be implemented on the independent subsets, tuning the networks' parameters such as depth and number of neurons in each layer, and the remaining data (15%) are devoted to monitoring the performance of NN during the training.

Table 7 presents the results of DEM fusion over individual target areas, employing different strategies of data selection for training. The results show nearly identical results for both strategies. Thus, pouring all subsets from different land types to make a general predictor model decreases the number of necessary ANNs from six to one for each kind of DEM. On the contrary, the size of the input data for the training becomes larger; thus, training requires more runtime. The runtime of NN training adopting different strategies (A and B), implemented in a system equipped with Intel(R) Core(TM) i7-6700, 3.40 GHz CPU and 16 GB RAM is collected in Table 6. The total runtime of strategy A for training both networks of Cartosat-1 and TanDEM-X is 158.1 s while training according to strategy B takes 519.6 s. This confirms that strategy A is computationally more economical for training.

In Table 7, we can also observe the RMSEs of fused DEMs generated by the weight maps that computed by adopting different weighting strategies. The results display slightly lower RMSE values by the $\frac{1}{e_i^2}$ weighting formulation for some areas like industrial, inner city and forested areas. On the other, the qualities of fused DEMs following the different training strategies (A and B) are almost same. The benefit of using strategy B is the establishment of a general predictor model that can be used for any arbitrary land types; while in strategy A, the networks have to be trained six times (for six different land types) and achieved predictor models must be used for respective land types in target areas, requiring semantic classification of the study area to identify different land types for DEM fusion.



(a) Slope (b) Aspect (c) ACV

(d) TRI (e) TPI (f) Roughness

(g) Ruggedness (h) SRF (i) Entropy

(j) Edginess

Fig. 5. Height error patterns of the TanDEM-X DEM (in red) and the Cartosat-1 DEM (in black dashed) for industrial area (Subset I1): horizontal directions show the feature values and vertical directions indicate mean absolute height residuals in each bin achieved from the refinement step. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A final experiment was implemented for comparing the DEM fusion results by employing ANN-based and HEM-based approaches with the initial DEMs. For this purpose, using ANNs with one hidden layer and 20 neurons in the hidden layer and the training based on the subsets D1, I1, H1, R1 and F1 (following strategy B as the economical way), the predictive models were created. At last, the full DEM fusion chain was carried out on totally independent subsets to evaluate the capability of the proposed algorithm for DEM fusion. These independent subsets were selected as target areas from different land types (subsets D2, I2, H2, R2, F2 and A). The TanDEM-X and Cartosat-1 DEMs of these subsets

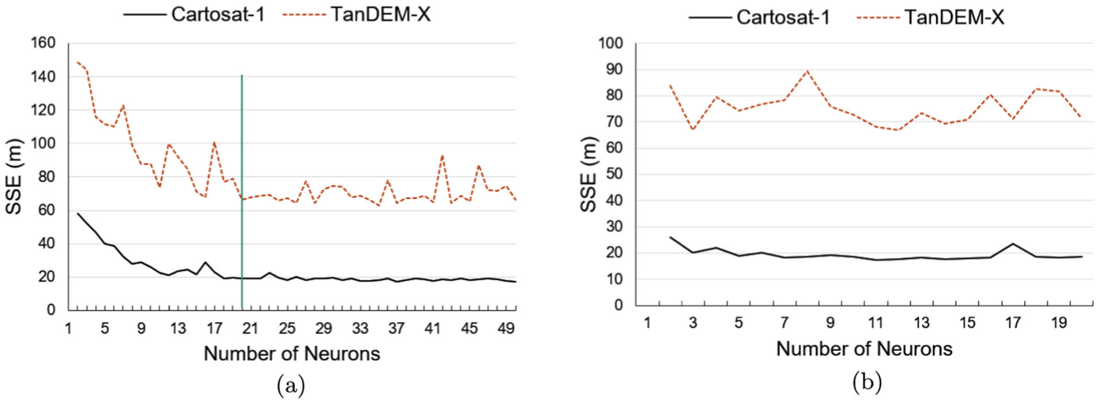


Fig. 6. The performance of the ANNs with different structures measured by SSE (Sum of Squares Error); (a) The structure with one hidden layer. (b) The structure that organized by two hidden layers: first, with number of neurons fixed to n = 20, and second with varying number of neurons.

were fused together by weighted averaging using the achieved height error maps as weights of the input DEMs by the $\frac{1}{e_i^2}$ weighting formulation. It has to be noted, as illustrated in Section 1, the quality of the TanDEM-X DEM is significantly worse than the Cartosat-1 DEM in the Lake subset, so that heights of the TanDEM-X DEM should be completely substituted by the heights of the Cartosat-1 DEM.

The results are summarized in Table 8 and compared to simple HEM-based fusion. Fig. 8 visualizes the absolute residual maps of TanDEM-X, Cartosat-1, HEM and ANN-based DEM fusion results in comparison to reference data for some exemplary study areas.

6. Discussion

As the obtained NMAD and RMSE results show, standard HEMs generally cannot be reliably used to produce a fused DEM whose height precision exceeds the Cartosat-1 DEM precision. This confirms the assumption that standard HEMs do not reflect all possible error sources in the original DEM data. As an example, the HEM delivered with the TanDEM-X raw DEM just contains error values derived from interferometric coherence and baseline configuration, while deterministic error sources such as layover are not considered. Nevertheless, standard HEMs can be used as a fall-back solution should ground truth for ANN-based weight map prediction be unavailable.

In contrast, the results obtained for the ANN-supported fusion shows an improvement of the fused DEM product with respect to both input datasets, indicating that the designed ANNs can properly model the existing error patterns related to spatial features that describe the landscaping and the roughness of the land surface under investigation. While the ANN-supported DEM fusion significantly improves the height precision of the TanDEM-X DEM, it also enhances the quality of the Cartosat-1 DEM: As an additional analysis reveals, more than 51% of all fused DEM pixels are more accurate than their Cartosat-1 counterparts.

As can be seen in Table 3, the size of training subsets is usually at least as large as the size of the corresponding target subsets, which could lead to the misconception that the number of training and test samples needs to be similar necessarily. In order to show that this is not the case, we conducted an exemplary experiment evaluating the impact of the training dataset size for the inner city subset. Fig. 9 illustrates the influence of the training sample number on the ANN-predicted weights. Since the actual DEM fusion results depend only on those weights, it can be seen that as few as about 2000 training samples already provide a stable weight prediction and thus stable fusion results. This is due to using a shallow ANN architecture with only a couple of neurons, which produces stable predictions already with a limited amount of training

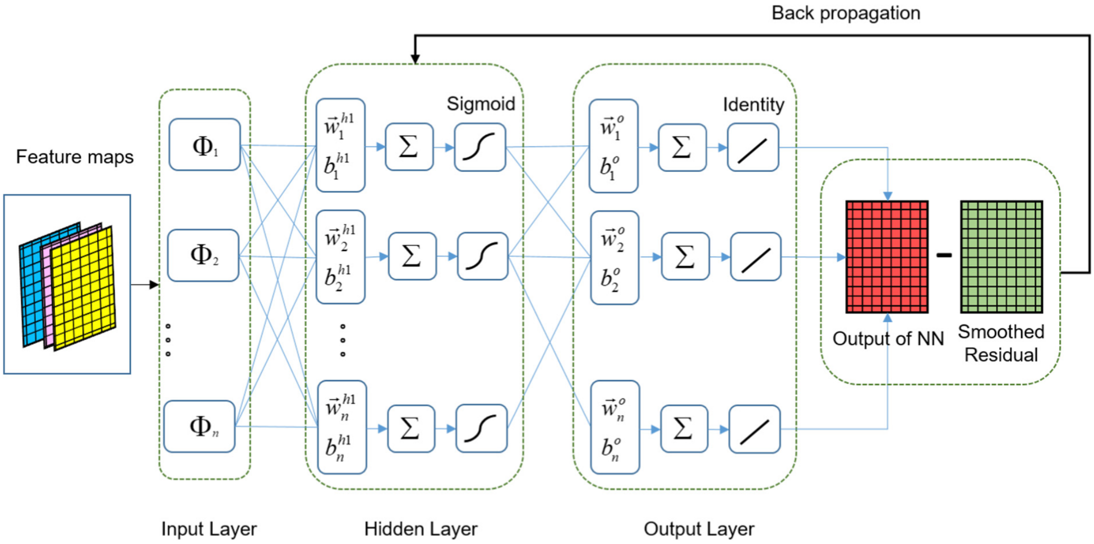


Fig. 7. Structure of neural network for weight map prediction.

Table 5
Height precision (in meters) of TanDEM-X DEM, the Cartosat-1 DEM and final fused DEM using the HEMs and STD of matching as weight maps in WA (in meters). The green shaded values indicate the best values of metrics relevant to the quality of the compared DEMs.

Areas		TanDEM-X		Cartosat-1		Fused DEM	
		NMAD	RMSE	NMAD	RMSE	NMAD	RMSE
Urban	Industrial: D1	2.68	4.61	1.66	3.24	1.95	3.57
	Industrial: D2	3.44	5.12	2.44	3.56	2.92	4.27
	Inner city: I1	5.57	6.43	4.34	5.27	4.64	5.33
	Inner city: I2	5.92	5.81	4.63	5.22	5.09	5.13
	High building: H1	6.72	18.17	3.63	13.10	4.69	16.12
	High building: H2	3.76	8.41	3.29	8.13	3.14	7.93
	Residential: R1	2.95	3.10	2.56	2.83	2.81	2.90
	Residential: R2	2.30	2.61	2.02	2.45	2.13	2.44
Non-Urban	Forested: F1	—	—	—	—	—	—
	Forested: F2	3.88	4.82	3.23	4.65	3.51	4.35
	Agricultural (F1)	—	—	—	—	—	—
	Agricultural (A)	0.93	0.84	0.65	0.81	0.98	0.76

Table 6
The training computational cost using different strategies, A and B.

Strategy	DEM	Subset	Runtime (second)
A	Cartosat-1	Industrial: D1	33.2
		Inner city: I1	10.8
		High building: H1	0.1
		Residential: R1	0.8
		Forested: F1	2.4
		Agricultural: F1	8.9
	TanDEM-X	Industrial: D1	59.7
		Inner city: I1	15.6
		High building: H1	0.1
		Residential: R1	1.4
		Forested: F1	6.8
		Agricultural: F1	18.3
B	Cartosat-1	All	187.5
	TanDEM-X	All	332.1

Table 7
Results of fusing TanDEM-X and Cartosat-1 DEM (in meters) by employing weight maps predicted by ANN with the application of different training strategies and different types of weighting.

	Fused DEM			
Training strategy	Individual	Individual	All	All
Weight	$\frac{1}{e_i^2}$	$1-e_{in}$	$\frac{1}{e_i^2}$	$1-e_{in}$
Areas	RMSE	RMSE	RMSE	RMSE
Industrial: D2	3.46	3.63	3.52	3.62
Inner city: I2	4.84	4.94	4.85	4.92
High building: H2	7.75	7.71	7.75	7.66
Residential: R2	2.33	2.38	2.34	2.35
Forested: F2	4.18	4.28	4.18	4.25
Agricultural: A	0.70	0.69	0.68	0.69

data. This is also supported by the preprocessing phase described in Section 4.1.2. The experiment confirms that the sizes of training and target areas need not be similar for our method to work. Nevertheless, for the preprocessing and feature extraction, as well as for providing validation data during the training, patch-shaped subsets should be selected instead of selecting only few independent DEM pixels.

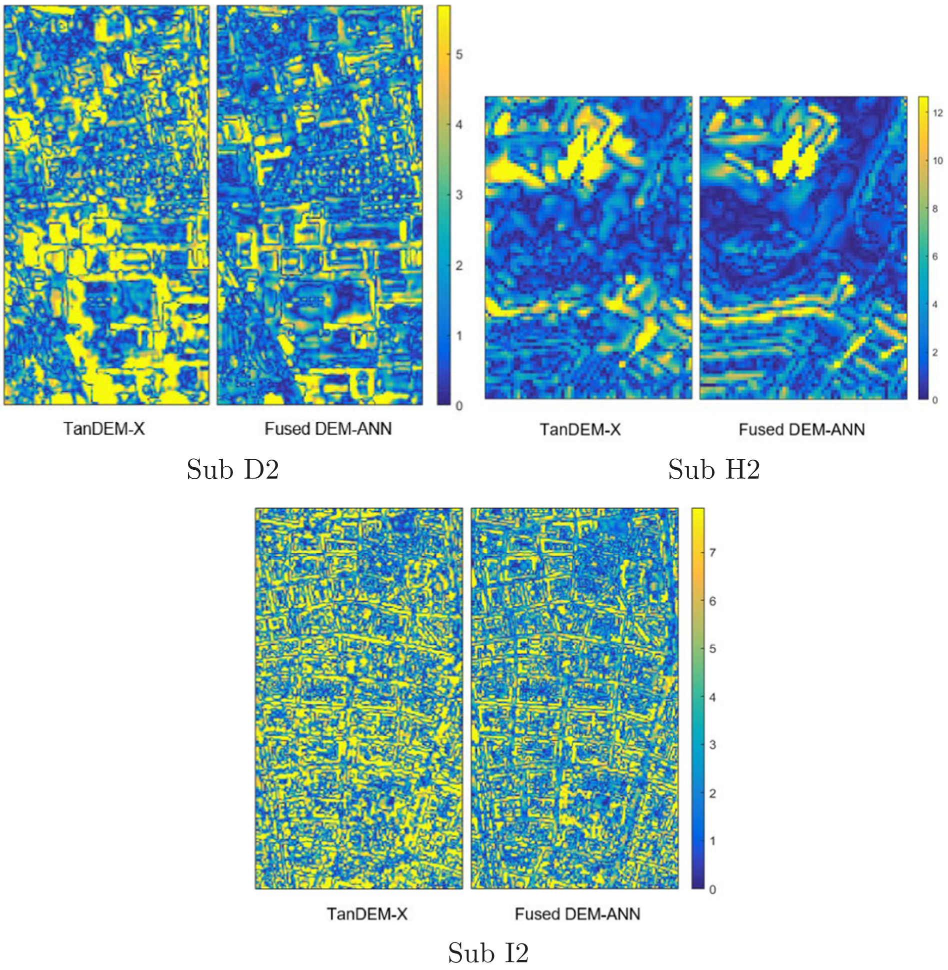
Last but not least, the absolute vertical accuracy of the fused DEM is also better than the absolute accuracy of the original Cartosat-1 DEM, which is achieved through the alignment to the more accurately localized TanDEM-X DEM. Thus, eventually, the proposed DEM fusion is able to provide a final DEM product that provides a higher quality than the individual input DEMs in both absolute and relative measures.

7. Evaluation of the ANN-based fusion algorithm using other DEM data

While the main focus of this study is to fuse the TanDEM-X and Cartosat-1 DEMs according to the reasons mentioned in Section 1, we also seek to validate the proposed framework for the fusion of other global DEM data. Thus, we carried out similar experiments for ASTER GDEM and SRTM-C DEMs over the Munich area. Fig. 10 displays both the training and target areas including the same different urban and sub-urban land types. SRTM data with pixel spacing of 1 arcsec (about 30 m) from the C-band data-takes by Shuttle Radar Topography Mission (operated in February 2000) are globally available by USGS portal as InSAR DEMs (USGS, 2000). In this experiment the void filled SRTM DEM was used. Moreover, version 002 of the ASTER GDEMs was produced from thermal ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) sensor (initially launched in 1999) by stereoscopic 3D reconstruction (USGS, 1999). In addition, the Cartosat-1 data are used as a ground truth both for purposes of height residual estimation for training of the ANN and assessing the final fusion chain.

Table 8
Results of fusion of TanDEM-X and Cartosat-1 DEM (in meters) by using weight maps generated by different methods. The weight maps generated by $\frac{1}{e_i^2}$ and strategy B (see Section 4.1.3) were finally adopted for training the ANNs. The green shaded values indicate the best values of metrics relevant to the quality of the compared DEMs.

Areas	TanDEM-X		Cartosat-1		Fused DEM			
					HEM		ANN	
	NMAD	RMSE	NMAD	RMSE	NMAD	RMSE	NMAD	RMSE
Industrial: D2	3.43	5.11	2.43	3.56	2.91	4.26	2.42	3.46
Inner city: I2	5.92	5.81	4.62	5.21	5.09	5.13	4.90	4.84
High building: H2	3.75	8.41	3.29	8.13	3.14	7.93	3.13	7.75
Residential: R2	2.30	2.61	2.02	2.45	2.13	2.44	1.99	2.33
Forested: F2	3.88	4.82	3.23	4.65	3.51	4.35	3.19	4.18
Agricultural: A	0.57	0.84	0.65	0.81	0.98	0.76	0.49	0.68



Sub D2

Sub H2

Sub I2

Fig. 8. Absolute residual maps of TanDEM-X and Fused DEM using ANN-predicted weight maps in some exemplary subsets.

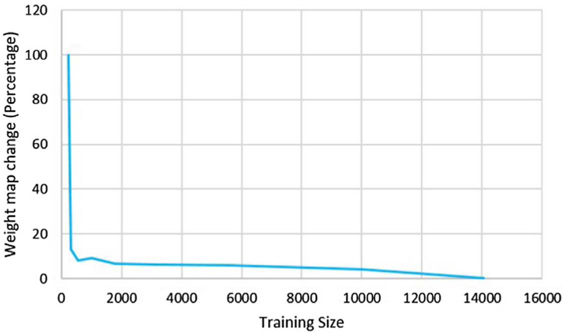


Fig. 9. The change of weight map (percentage) according to change in size of training data.

As we illustrated in Section 5, it is not so necessary to take care of strategies A or B for training or different ways of weighting, but for making the process of DEM fusion easier, following strategy B can cope with the task of DEM fusion. As a result, ANN-based fusion according to strategy B was employed for the new DEM data fusion. The fusion results are expressed in Table 9. Because no weight maps were provided with the elevation data, the simple averaging approach was used instead of WA for comparison of the results with our proposed method. The results confirm the efficiency of the proposed algorithm for new cases of DEM fusion as well. It can be clearly inferred from the results that simple averaging just improves the accuracy of the ASTER DEM, and RMSE value of fused DEM does not exceed the RMSE of SRTM DEM, while by using the ANN and following the invented framework in this study, the final RMSE of fused DEM becomes better than initial study DEMs—meaning the more reliable DEM can be obtained by NNs. In terms of NMAD metric, the quality of ANN-based DEM fusion is significantly higher than simple averaging.

8. Conclusion

This study focused on the fusion of TanDEM-X and Cartosat-1 DEM over urban areas and their surroundings. The main objective of the investigation was to ultimately obtain the final fused DEM with higher height precision and absolute accuracy. For this task, a typical data fusion technique, weighted averaging, is employed to pixel-wise fuse the elevation data. To achieve optimal fusion results, an innovative framework was developed to predict weight maps using a fully connected artificial neural network. The results demonstrated that the proposed method can efficiently improve the height precision of both Cartosat-1 and TanDEM-X DEMs up to 50% in urban areas and 22% in non-urban areas as well as the absolute accuracy could be increased through the data alignment. While the height precision improvement by the HEM-based method does not exceed 20% and 10% in urban and

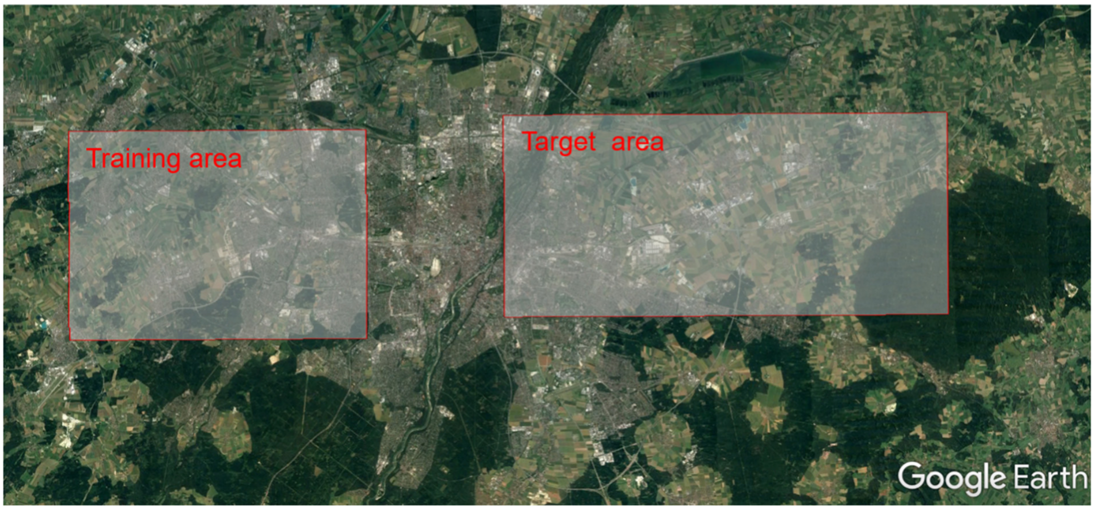


Fig. 10. Location of training and target areas for ASTER and SRTM DEM fusion study.

Table 9
ASTER GDEM and SRTM-C data fusion results over Munich area using the proposed NN-based approach.

	SRTM-C		ASTER GDEM		Fused DEM			
	Original		Original		Averaging		ANN	
	NMAD	RMSE	NMAD	RMSE	NMAD	RMSE	NMAD	RMSE
Target area	2.25	4.01	4.92	6.51	3.14	4.51	2.34	3.89

non-urban areas respectively if HEMs would be available for DEM fusion. Finally, for validating the application of the NN-based approach in other cases of InSAR and optical DEM fusions, different data such as the ASTER GDEM and the SRTM-C elevation data were fused in this way. The results again proved the efficiency of the proposed algorithm for optical and InSAR DEM fusion applications.

Acknowledgment

The authors would like to thank Prof. Reinartz and Dr. d́Angelo of DLR for providing the Cartosat-1 DEM; Dr. Fritz of DLR for providing the TanDEM-X raw DEM; and the Bavarian Surveying Administration for providing the LiDAR reference data. This work is jointly supported by the European Research Council (ERC) under the European Union Horizon 2020 research and innovation program (grant agreement No [ERC-2016-StG-714087], Acronym: *So2Sat*), and Helmholtz Association under the framework of the Young Investigators Group “SiPEO” (VH-NG-1018, www.sipeo.bgu.tum.de).

References

Ahmed, N., Mahtab, A., Agrawal, R., Jayaprasad, P., Pathan, S.K., Ajai, Singh, D.K., Singh, A.K., 2007. Extraction and validation of Cartosat-1 DEM. J. Indian Soc. Remote Sens. 35, 121.

Bagheri, H., Schmitt, M., Zhu, X.X., 2017. Uncertainty assessment and weight map generation for efficient fusion of TanDEM-X and Cartosat-1 DEMs. In: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-1/W1, pp. 433–439.

Bagheri, H., Schmitt, M., Zhu, X.X., 2017. Fusion of TanDEM-X and Cartosat-1 DEMs using TV-norm regularization and ANN-predicted weights. In: Proceedings of IEEE Geosci. and Remote Sens. Symposium.

Birgé, L., Rozenholc, Y., 2006. How many bins should be put in a regular histogram. ESAIM: Probab. Stat. 10, 24–45.

d’Angelo, P., Lehner, M., Krauss, T., Hoja, D., Reinartz, P., 2008. Towards automated DEM generation from high resolution stereo satellite images. In: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVII, pp. 1137–1342.

Deo, R., Rossi, C., Eineder, M., Fritz, T., Rao, Y.S., 2015. Framework for fusion of ascending and descending pass TanDEM-X raw dems. IEEE J. Sel. Top. Appl. Earth Obser. Remote Sens. 8, 3347–3355.

Fritz, T., Rossi, C., Yague-Martinez, N., Rodriguez-Gonzalez, F., Lachaise, M., Breit, H., 2011. Interferometric processing of TanDEM-X data. In: 2011 IEEE International Geoscience and Remote Sensing Symposium, pp. 2428–2431.

Fuss, C.E., Berg, A.A., Lindsay, J.B., 2016. Dem fusion using a modified k-means clustering algorithm. Int. J. Digital Earth 9, 1242–1255.

Gruber, A., Wessel, B., Huber, M., Roth, A., 2012. Operational TanDEM-X DEM calibration and first validation results. ISPRS J. Photogramm. Remote Sens. 73, 39–49 (Innovative Applications of SAR Interferometry from modern Satellite Sensors).

Gruber, A., Wessel, B., Martone, M., Roth, A., 2016. The TanDEM-X DEM mosaicking: fusion of multiple acquisitions using InSAR quality parameters. IEEE J. Sel. Top. Appl. Earth Obser. Remote Sens. 9, 1047–1057.

Heady, B., Kroenung, G., Rodarmel, C., 2009. High resolution elevation data (HRE) specification overview. In: ASPRS/MAPPS 2009 Conference, San Antonio, Texas.

Höhle, J., Höhle, M., 2009. Accuracy assessment of digital elevation models by means of robust statistical methods. ISPRS J. Photogramm. Remote Sens. 64, 398–406.

Just, D., Bamler, R., 1994. Phase statistics of interferograms with applications to synthetic aperture radar. Appl. Opt. 33, 4361–4368.

Kim, T., Jeong, J., 2011. DEM matching for bias compensation of rigorous pushbroom sensor models. ISPRS J. Photogramm. Remote Sens. 66, 692–699.

Krieger, G., Moreira, A., Fiedler, H., Hajnsek, I., Werner, M., Younis, M., Zink, M., 2007. TanDEM-X: a satellite formation for high-resolution SAR interferometry. IEEE Trans. Geosci. Remote Sens. 45, 3317–3341.

Kuschk, G., d’Angelo, P., Gaudrie, D., Reinartz, P., Cremers, D., 2017. Spatially regularized fusion of multiresolution digital surface models. IEEE Trans. Geosci. Remote Sens. 55, 1477–1488.

Lehner, M., Müller, R., Reinartz, P., Schroeder, M., 2007. Stereo evaluation of Cartosat-1 data for French and Catalonian test sites. In: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI-1/W51 (on CD–ROM).

Martone, M., Bräutigam, B., Rizzoli, P., Gonzalez, C., Bachmann, M., Krieger, G., 2012. Coherence evaluation of TanDEM-X interferometric data. ISPRS J. Photogramm. Remote Sens. 73, 21–29.

Olaya, V., 2009. Chapter 6: Basic land-surface parameters. In: T. H. Science, H. I. R. B. T. D. in Soil (Eds.), GeomorphometryConcepts, Software, Applications, vol. 33. Elsevier, pp. 141–169.

Papasaika, H., Poli, D., Baltsavias, E., 2009. Fusion of digital elevation models from various data sources. In: 2009 International Conference on Advanced Geographic Information Systems Web Services, pp. 117–122.

Papasaika, H., Kokiopoulou, E., Baltsavias, E., Schindler, K., Kressner, D., 2011. Fusion of digital elevation models using sparse representations. In: Proceedings of the 2011 ISPRS Conference on Photogrammetric Image Analysis, PIA’11. Springer-Verlag, Berlin, Heidelberg, pp. 171–184.

Pock, T., Zebedin, L., Bischof, H., 2011. TGV-Fusion. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 245–258.

Rabus, B., Eineder, M., Roth, A., Bamler, R., 2003. The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborne radar. ISPRS J. Photogramm. Remote Sens. 57, 241–262.

Ravanbakhsh, M., Fraser, C.S., 2013. A comparative study of DEM registration approaches. J. Spat. Sci. 58, 79–89.

Reinartz, P., Müller, R., Hoja, D., Lehner, M., Schroeder, M., 2005. Comparison and fusion of DEM derived from SPOT-5 HRS and SRTM data and estimation of forest heights. In:

Earsel (Ed.), Earsel Symposium, Porto, Portugal, 6.–11. June 2005.

Reinartz, P., d'Angelo, P., Krauß, T., Poli, D., Jacobsen, K., Buyuksalih, G., 2010. Benchmarking and quality analysis of DEM generated from high and very high resolution optical stereo satellite data. In: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII.

Rodriguez, E., Morris, C.S., Belz, J.E., 2006. A global assessment of the SRTM performance. Photogramm. Eng. Remote Sens. 72, 249–260.

Rossi, C., Gernhardt, S., 2013a. Urban DEM generation, analysis and enhancements using TanDEM-X. ISPRS J. Photogramm. Remote Sens. 85, 120–131.

Rossi, C., Fritz, T., Breit, H., Eineder, M., 2011. First urban TanDEM-X raw DEMs analysis. In: 2011 Joint Urban Remote Sensing Event, pp. 65–68.

Rossi, C., Gonzalez, F.R., Fritz, T., Yague-Martinez, N., Eineder, M., 2012. TanDEM-X calibrated raw DEM generation. ISPRS J. Photogramm. Remote Sens. 73, 12–20.

Rossi, C., Eineder, M., Fritz, T., d'Angelo, P., Reinartz, P., 2013. Quality assessment of TanDEM-X raw DEMs oriented to a fusion with CartoSAT-1 DEMs. In: 33rd EARSeL Symposium, pp. 1–9.

Roth, A., Knopfle, W., Strunz, G., Lehner, M., Reinartz, P., 2002. Towards a global elevation product: combination of multi-source digital elevation models. Int. Arch. Photogramm. Remote Sens. Spat. Inform. Sci. 34, 675–679.

Schmitt, M., Zhu, X.X., 2016. Data fusion and remote sensing: an ever-growing relationship. IEEE Geosci. Remote Sens. Mag. 4, 6–23.

Srivastava, P.K., Srinivasan, T., Gupta, A., Singh, S., Nain, J.S., Prakash, S., Kartikeyan, B., Krishna, B.G., 2007. Recent advances in CARTOSAT-1 data processing. In: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI-1/W51 (on CD–ROM).

Tachikawa, T., Kaku, M., Iwasaki, A., Gesch, D.B., Oimoen, M.J., Zhang, Z., Danielson, J.J., Krieger, T., Curtis, B., Haase, J., et al., 2011. ASTER Global Digital Elevation Model Version 2-summary of Validation Results. Technical Report. NASA.

Tadono, T., Ishida, H., Oda, F., Naito, S., Minakawa, K., Iwamoto, H., 2014. Precise global DEM generation by ALOS PRISM. ISPRS Ann. Photogram., Remote Sens. Spat. Inform. Sci. 2, 71.

Takaku, J., Tadono, T., Tsutsui, K., 2014. Generation of high resolution global DSM from ALOS PRISM. Int. Arch. Photogram., Remote Sens. Spat. Inform. Sci. 40, 243.

Takaku, J., Tadono, T., Tsutsui, K., Ichikawa, M., 2016. Validation ofAW3D global DSM generated from ALOS PRISM. ISPRS Ann. Photogram., Remote Sens. Spat. Inform. Sci. 3, 25.

The Federal Agency for Cartography and Geodesy of Germany (BKG), 2016. Digital orthophotos. < https://www.bkg.bund.de/SharedDocs/Downloads/BKG/DE/Downloads-DE-Flyer/AdV-DOP-DE > (accessed 09.17).

Teo, T.-A., 2011. Bias compensation in a rigorous sensor model and rational function model for high-resolution satellite images. Photogramm. Eng. Remote Sens. 77, 1211–1220.

Toutin, T., 2002. Impact of terrain slope and aspect on radargrammetric DEM accuracy. ISPRS J. Photogramm. Remote Sens. 57, 228–240.

USGS, 1999. Routine ASTER Global Digital Elevation Model. < https://lpdaac.usgs.gov/dataset_discovery/aster/aster_products_table/astgtm > .

USGS, 2000. Shuttle Radar Topography Mission (SRTM) Void Filled. < https://lta.cr.usgs.gov/SRTMVF > (accessed 09.17).

Uttenthaler, A., Barner, F., Hass, T., Makiola, J., d'Angelo, P., Reinartz, P., Carl, S., Steiner, K., 2013. EURO-MAPS 3D- a transnational, high-resolution digital surface model for Europe. In: ESA Living Planet Symposium, vol. 722, p. 271.

Zink, M., Bachmann, M., Brautigam, B., Fritz, T., Hajnsek, I., Moreira, A., Wessel, B., Krieger, G., 2014. TanDEM-X: the new global DEM takes shape. IEEE Geosci. Remote Sens. Mag. 2, 8–23.

A.2 H. Bagheri, M. Schmitt, and X. X. Zhu. Fusion of urban TanDEM-X raw DEMs using variational models. In: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11.12 (2018), pp. 4761–4774.

Fusion of Urban TanDEM-X Raw DEMs Using Variational Models

Hossein Bagheri, Michael Schmitt, *Senior Member, IEEE*, and Xiao Xiang Zhu, *Senior Member, IEEE*

***Abstract*—Recently, a new global digital elevation model (DEM) with pixel spacing of 0.4 arcsec and relative height accuracy finer than 2 m for flat areas (slopes $< 20\%$) and better than 4 m for rugged terrain (slopes $> 20\%$) was created through the TanDEM-X mission. One important step of the chain of global DEM generation is to mosaic and fuse multiple raw DEM tiles to reach the target height accuracy. Currently, weighted averaging (WA) is applied as a fast and simple method for TanDEM-X raw DEM fusion, in which the weights are computed from height error maps delivered from the Integrated TanDEM-X Processor (ITP). However, evaluations show that WA is not the perfect DEM fusion method for urban areas, especially in confrontation with edges such as building outlines. The main focus of this paper is to investigate more advanced variational approaches such as TV-L_1 and Huber models. Furthermore, we also assess the performance of variational models for fusing raw DEMs produced from data takes with different baseline configurations and height of ambiguities. The results illustrate the high efficiency of variational models for TanDEM-X raw DEM fusion in comparison to WA. Using variational models could improve the DEM quality by up to 2 m, particularly in inner city subsets.**

***Index Terms*—Data fusion, Huber model, L_1 norm total variation, TanDEM-X DEM, weight map.**

I. Introduction

GLOBAL digital elevation models (DEMs) with large coverage of the landmasses are an important source of geoinformation for different applications such as environmental studies, geographic information systems, remote sensing, etc. SAR interferometry is one of the main techniques being employed for global DEM productions because of its capability to cover large areas independent of daylight or weather. For example, a global DEM with coverage of most of the planet (between 56 °S and 60 °N) was generated by Shuttle Radar Topography Mission (SRTM). The SRTM DEM is provided in the form of tiles with pixel spacings of 1 arcsec ($\sim$30 m) and 3 arcsec ($\sim$90 m), respectively [1].

Recently, a new global DEM with even higher resolution (namely a pixel spacing of 0.4 arcsec) covering almost the whole planet was realized by the TanDEM-X DEM mission. Again, bistatic SAR acquisitions are used as input to a SAR interferometric processing chain to produce the DEM. The primary target of the mission was to provide global DEM with relative height accuracy better than 2 m for flat areas (slopes lower than 20%) and finer than 4 m for remaining steeper slopes [2]. For this, bistatic SAR data serve as an excellent data source, reducing atmospheric effects and avoiding temporal decorrelation in the InSAR process. Form raw SAR data take to the final global DEM, a workflow including different phases such as interferogram generation, phase unwrapping (PU), data calibration, DEM block adjustment, and mosaicking is implemented at DLR [3]. A main step of the DEM generation procedure is carried out in the Integrated TanDEM-X Processor (ITP), which leads to primary raw DEMs for each bistatic acquisition [4]. During the raw DEM generation, some potential error sources are removed by instrument and baseline calibration [5]. After that, the vertical bias, which usually lies between 1 and 5 m, is corrected by a least squares block adjustment [6]. The block adjustment is performed by using ICESat data and connecting points in the overlapping areas of raw DEM tiles. However, dependent on the terrain morphology, some error sources still remain after the block adjustment. The effect of these errors can be decreased through fusion of several DEM coverages within the DEM Mosaicking Processor (DMP) [7]. The TanDEM-X raw DEM coverage over different terrain types is displayed in Fig. 1. As can be seen, the most of the world is covered by at least two nominal acquisitions with height of ambiguities (HoA) between 30 and 55 m. The main objective of TanDEM-X DEM fusion is to improve the final accuracy by employing several coverages over different areas [8].

Diverse methods have been designed for the fusion DEMs with different properties, which can be seen as an application of data fusion in remote sensing [9]. Among them, weighted averaging (WA) is well established as a simple approach with low computational cost [8], [10]–[12]. However, its performance strongly depends on the weights that describe the height error distribution for each pixel [13]. For SAR interferometric-derived DEMs, the weights can be achieved from height error maps (HEM) [14], which are a byproduct of the InSAR process and derived from the coherence values and the given geometrical configuration [15]. However, it should be noted that HEMs

Manuscript received June 20, 2018; revised September 6, 2018; accepted October 17, 2018. Date of publication November 18, 2018; date of current version December 31, 2018. The work of X. Zhu is jointly supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under Grant ERC-2016-StG-714087 (Acronym: *So2Sat*), Helmholtz Association under the framework of the Young Investigators Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de), and the Bavarian Academy of Sciences and Humanities in the framework of Junges Kolleg. *(Corresponding author: Xiaoxiang Zhu.)*

H. Bagheri and M. Schmitt are with the Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich 80333, Germany (e-mail: hossein.bagheri@tum.de; m.schmitt@tum.de).

X. Zhu is with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Cologne 51147, Germany, and also with the Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich 80333, Germany (e-mail: xiaoxiang.zhu@dlr.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2018.2878608

1939-1404 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

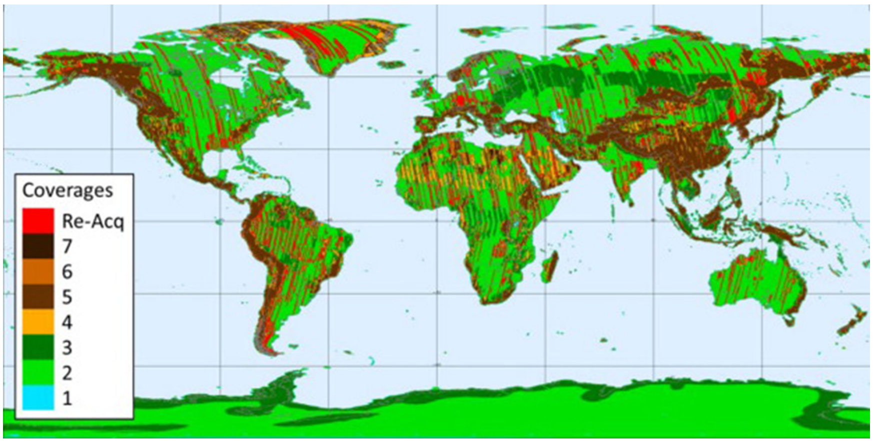


Fig. 1. TanDEM-X coverage in different areas [3].

cannot represent all error sources as they do not reflect deterministic effects such as layover and shadow effects. Another way is to compute weight maps by a comparison with ground truth data that are not necessarily available for every arbitrary study area [16].

A current approach for implementing the DEM fusion in DMP is WA. In addition to WA, some logic for clustering consistent heights and upgrading weights regarding the influences of other significant factors such as HoA, PU methodology, and pixel locations relative to the border of the DEM scene is considered to finally reach the target relative accuracy and minimize PU errors remaining from primary steps [8]. While the WA approach can realize the predefined goals in the DMP for global DEM generation, it does not perform optimally in difficult terrains with complex morphology such as urban areas, which contains many high-frequency contents such as edges. After WA-based TanDEM-X DEM fusion, visualization shows that outlines of buildings are not perfectly sharp and still some amount of existing noise spoils the footprints of buildings [for example, see Fig. 9(c)]. As Fig. 1 illustrates, most areas are only covered by two nominal acquisitions (shown in green). Since this holds for many important urban areas as well, this also motivates the development of more sophisticated approaches.

An advanced approach for DEM fusion was proposed by Papasaika *et al.* [17]. They exploited sparse representations for multisensor DEM fusion. Zach *et al.* implemented an L_1 norm total variational model for range image fusion [18]. In another study, Pock *et al.* proposed to use total generalized variation (TGV) for fusing DEMs derived from airborne optical imagery [19]. Kuschk *et al.* evaluated weighted TGV to fuse DEMs derived from spaceborne optical imagery with different resolutions [20]. In another study, weighted TV-L_1, in which weights were predicted by neural networks, were applied for the Cartosat-1 and TandEM-X DEM fusion over urban areas [21]. Overall, in spite of the high computational cost of advanced methods for DEM fusion, they perform more efficiently than simple WA.

In this paper, we will investigate the application of more sophisticated DEM fusion approaches that are able to efficiently preserve edges and outlines of buildings while still reducing noise effects. For this purpose, two variational models, namely L_1 norm total variation (TV-L_1) and Huber model are implemented. Apart from these regularization approaches, we will also investigate the potential of employing raw DEMs with different properties such as different baseline configuration for the TanDEM-X DEM fusion. Therefore, this paper is structured into several sections. In Section II, the methodology of DEM fusion based on regularization methods is explained. Then, the description of the study subsets and experimental results from DEM fusion are provided in Section III. Finally, the performance of the implemented DEM fusion methods for TanDEM-X data over urban areas will be discussed in Section IV.

II. Methods for TanDEM-X Raw DEM Fusion

In this paper, two approaches are implemented for TanDEM-X raw DEM fusion. The characteristics of each model will be explained in the following. Before the fusion, raw DEMs at first are aligned to each other by DEM coregistration approaches such as least squares matching [22], iterative closest point [23], or manual registration. The coregistration of DEMs decreases their translational and rotational differences. For stability reasons, in addition, the height data should be normalized to the interval [0, 1] [20]

$$h_k^n(x,y) = \frac{h_k(x,y) - h_{\text{min}}}{h_{\text{max}} - h_{\text{min}}} \tag{1}$$

where $h_k(x,y) > 0$ is the elevation of the study DEM with index k at location (x,y), $h_{\text{max}} > 0$ and $h_{\text{min}} > 0$ ($h_{\text{min}} < h_{\text{max}}$) are the lowest and highest elevations among all input DEMs. The output gives the normalized height in the considered location.

A. Background: Weighted Averaging

The most popular, very fast, and low computational cost method for DEM fusion is WA, which is implemented by

$$\mathbf{f} = \sum_{i=1}^{k} \mathbf{w}_i \odot \mathbf{h}_i \tag{2}$$

where $\mathbf{h}_i$ are 2-D arrays representing the input DEMs, $\mathbf{w}_i$ are the corresponding weight maps, and $\odot$ is a pixelwise product. It is worth to note that other simple methods, such as a pixelwise median- or mode-based fusion, can also be employed for DEM fusion, especially when multiple DEMs are available [24].

As explained in Section I, the main critical issue for using WA for DEM fusion is to apply appropriate weights that are fairly representative of expected height errors in the source DEMs. For TanDEM-X DEM fusion, generally, these weights are delivered as HEMs from the ITP. For each height of the TanDEM-X DEM, the corresponding HEM value can be estimated by

$$\sigma_j = H_{\text{amb}} \, \frac{\sigma_{\phi,j}}{2\pi} \tag{3}$$

where H_{amb} is the height of ambiguity and $\sigma_{\phi,j}$ is the interferometric phase error that is estimated from the interferometric coherence and the InSAR geometry [2]. Then, from these values, the respective weights can be calculated for each pixel location by

$$w_j = \frac{\frac{1}{\sigma_j^2}}{\sum_{j=1}^{N} \frac{1}{\sigma_j^2}}. \tag{4}$$

B. Regularization-Based Models

Variational models were first used for signal and image denoising [25], [26]. Generally, in variational denoising approaches, an energy functional is constituted by fidelity and regularization terms. The fidelity is considered to enforce the output image being similar to the input images while the regularization term (also called penalty term) is embedded to reduce the effect of noise in the final result. The desired output is achieved by minimizing the constructed energy functional. Diverse energy functionals can be formed according to different functions for defining data and penalty terms [19].

A popular type of variational models is the total variation-based model (TV) in which the gradient of a desired output image is selected to form the regularization term based on different norms. The main advantage of the TV-based variational model is its convexity that guarantees to find a solution by minimizing the energy functional.

In the problem of TanDEM-X DEM fusion, several input raw DEMs are fused using variational models. The data term makes the fused DEM similar to the input tiles while the TV-based regularization term is defined to provide a sharp output at the end by preserving the edges and reducing the noise. This property is beneficial for fusing TanDEM-X raw DEMs over urban areas where footprints of buildings as edges often appear very noisy because of the inherent SAR imaging properties.

The basic gradient-based variational model for image denoising and data fusion is a quadratic model in which L_2 norm is used for both regularization and data terms [27]. However, the quadratic regularization term causes oversmoothing for edges. Therefore, using the L_1 norm instead was proposed by Rudin *et al.* which is called ROF model correspondingly [25]. Since the ROF model still uses the L_2 norm for the data term, it does not provide robustness against outliers when applied to DEM fusion. As a solution, the L_1 norm can be substituted for the L_2 norm [28]. The TV-L_1 model consists of the data fidelity and the penalty term

$$\min_{\mathbf{f}} \left\{ \sum_{i=1}^{k} \|\mathbf{f} - \mathbf{h}_i\|_1 + \gamma \|\nabla \mathbf{f}\|_1 \right\} \tag{5}$$

where $\mathbf{h}_i$ are noisy input DEMs and $\mathbf{f}$ is the desired DEM, which should be achieved by minimizing the functional energy mentioned above. The penalty term is formed based on the gradients of the newly estimated DEM to preserve the edges at the end. The regularization parameter γ trades off between penalty and fidelity terms. Increasing γ will influence the smoothness and will produce a smoother fused DEM in the end.

While the main advantage of TV-L_1 is its robustness against strong outliers as well as edge preservation [19], it suffers from the staircasing effect, a phenomenon that creates artificial discontinuities in the final output and particularly affects high resolution DEM fusion [29]. Moreover, the L_1 norm is not necessarily the best choice for all data fusion and denoising cases. As an alternative, the Huber regularization model is proposed to rectify the drawbacks of the TV-L_1 model [19]. It applies to the Huber norm instead of the L_1 norm in both fidelity and penalty terms [30]

$$\|x\|_\eta = \begin{cases} \frac{|x|^2}{2\eta} & \text{if } |x| \le \eta \\ |x| - \frac{\eta}{2} & \text{if } |x| > \eta. \end{cases} \tag{6}$$

Here, η is a parameter that determines a threshold between the L_1 and L_2 norm in the model. Based on this, the Huber model can be defined as [31]

$$\min_{\mathbf{f}} \left\{ \sum_{i=1}^{k} \sum_{\Omega} \|\mathbf{f} - \mathbf{h}_i\|_\alpha + \gamma \sum_{\Omega} \|\nabla \mathbf{f}\|_\beta \right\} \tag{7}$$

where both data and penalty terms are constituted based on the thresholds α and β that are substituted as η in the Huber norm relation (6) to form these terms and Ω denotes the raster DEM space. It should be noted that the Huber norm is a generalized form of the L_1 norm. However, in this study, the Huber norm is also used to strictly penalize the outliers.

Using the quadratic norm in the regularization term penalizes high-frequency changes more than L_1 norm, and thus, it reduces the noise at the cost of oversmoothing edges. The Huber norm, dependent on η values, treats as a norm between L_1 and L_2 norms. However, for $\eta = 1$, its behavior is nearly similar to L_1. In other words, the Huber norm with higher η provides DEMs with smoother building footprints but not as much as the quadratic norm. The influence of the parameters of variational models on the quality of fused DEM will be discussed in Section IV-A with more details. In the remainder of this paper, we use $\alpha = 4$ to smooth relative height errors larger than 4 m (considering the relative accuracy of the TanDEM-X DEM), and $\beta = 1$ based on data-driven experiments on different datasets. Consequently, γ can be calculated by the L-curve method [32].

C. Implementation

It is mathematically proven that the TV-based energy functional based on L_1 or the Huber norm is convex. The main

Algorithm 1: Dual primal algorithm.

```
Input: Primary DEMs to configure primal problem
Initialization: $\tau\sigma\|K\| \leq 1, (\hat{\mathbf{u}}^0, \mathbf{v}^0) \in \mathbf{U} \times \mathbf{V}$,
    $\hat{\mathbf{u}}^0 = \mathbf{u}^0, \theta \in [0, 1]$,
for $i = 0$ to stopping criteria do
    $\mathbf{v}^{i+1} = (I + \sigma\partial\mathcal{F}^*)^{-1}(\mathbf{v}^i + \sigma K\hat{\mathbf{u}}^i)$
    $\mathbf{u}^{i+1} = (I + \tau\partial\mathcal{G})^{-1}(\mathbf{u}^i - \tau K^T\mathbf{v}^{i+1})$
    $\hat{\mathbf{u}}^{i+1} = \mathbf{u}^{i+1} + \theta(\mathbf{u}^{i+1} - \mathbf{u}^i)$
end for
Output: $\mathbf{u}$
```

characteristic of a convex problem is that the desired output (i.e., the global minimum) will be certainly found through an optimization process. One popular strategy for finding the minimum of a convex optimization is to reformulate the functional energy as a primal-dual problem [33]. For variational models, the energy functional can be expressed in a general form such as

$$\min_{\mathbf{u}} \{\mathcal{G}(\mathbf{u}) + \mathcal{F}(K\mathbf{u})\} \tag{8}$$

where $\mathcal{G}(\mathbf{u})$ is the data term, and $\mathcal{F}(K\mathbf{u})$ is the regularization term. K refers to an operator that is used for defining the regularization term (for TV-based variational models, this is the gradient ∇ of the desired output). In the TanDEM-X raw DEM fusion, $\mathbf{u}$ (as primal variable) is the final fused DEM ($\mathbf{f}$) and the energy functional (terms $\mathcal{G}(\mathbf{u})$, $\mathcal{F}(K\mathbf{u})$) is defined by relations (5) and (7). Then, the dual-problem formulation of the energy functional can be written as

$$\min_{\mathbf{u}} \max_{\mathbf{v}} \{\mathcal{G}(\mathbf{u}) + \langle \mathbf{v}, K\mathbf{u} \rangle - \mathcal{F}^*(\mathbf{v})\} \tag{9}$$

$$\text{where} \quad \mathcal{F}^*(\mathbf{v}^*) = \sup_{\mathbf{v} \in \mathbf{V}} \langle \mathbf{v}^*, \mathbf{v} \rangle - \mathcal{F}(\mathbf{v}) \tag{10}$$

and $\mathbf{v}$ is the dual variable and $\mathcal{F}^*$ is defined as the convex conjugate of $\mathcal{F}$. The dual-problem algorithm for minimizing (9) is presented in Algorithm 1. It should be noted that the median-based fusion of the input DEMs can be used to initialize $\mathbf{u}^0$ to speed up the optimization. More details of the algorithm can be found in [33].

III. Experiments

In this paper, we investigate TanDEM-X raw DEM fusion over urban areas by using diverse TV-based variational models such as TV-L_1 and Huber models. In addition, the effect of fusing raw DEMs with different baseline configurations will be investigated. The baseline configurations for TanDEM-X data differ by changing orbit direction (ascending or descending) and also changing HoA values. Furthermore, the results of DEM fusion implementation will be evaluated for different land types with an emphasize on urban areas. These land types are as follows.

1) Industrial areas that are characterized as areas with large buildings, often not very high.
2) Inner city areas that include very densely packed buildings, relatively high.
3) Residential areas that are typically specified with low-rise single family homes.

In addition, we also considered some nonurban study areas such as agricultural and forested areas to evaluate the performance of variational models in those areas too.

In TanDEM-X raw DEMs produced with a pixel spacing of 0.2 arcsec (around 6 m), most building footprints in industrial and inner city areas can be visualized but the height accuracy and quality of building shapes suffer from noise and systematic errors such as layover and shadow. The visualization and quality of building become worse in residential areas because of the small sizes and heights of buildings in these areas. However, we will evaluate the performance of variational models for enhancing the quality of buildings appeared in the final fused DEM over different aforementioned land types. After resampling and coregistration, the raw DEMs are fused by the different approaches explained in Section II.

TABLE I
Properties of the Nominal TanDEM-X Raw DEM Tiles for Munich Area

TanDEM-X raws DEMs: Munich area		
Acquisition Id	1023491	1145180
Acquisition mode	Stripmap	Stripmap
Center incidence angle	38.25°	37.03°
Equator crossing direction	Ascending	Ascending
Look direction	Right	Right
Polarization	HH	HH
Height of ambiguity	45.81m	53.21 m
Pixel spacing	0.2 arcsec	0.2 arcsec
HEM mean	1.33 m	1.58

A. *Fusion of TanDEM-X Raw DEMs With Similar Baseline Configuration*

Most of the global coverage achieved with the TanDEM-X raw DEMs is generated by two nominal bistatistic acquisition (see Fig. 1), but there are more tiles in overlapping areas at the border of the tiles. The first investigation includes data takes that have similar baseline configurations as well as HoAs. The study subsets are selected from two nominal TanDEM-X raw DEMs over Munich city in Germany. The characteristics of these raw DEMs are presented in Table I.

Fig. 2 displays the raw DEM tiles used for this experiment. From those, four subsets as representatives of different land types are extracted for the DEM fusion task. A display of these subsets is provided in Fig. 3.

The quality of the input raw TanDEM-X DEMs as well as the fused version are determined by using the reference LiDAR DEM, which is produced from a high resolution airborne LiDAR point cloud acquired over Munich and provided by Bavarian Surveying Administration. The density of the LiDAR point cloud changes for each subset, but at least there is one point per square, and the vertical accuracy of the point cloud is better than

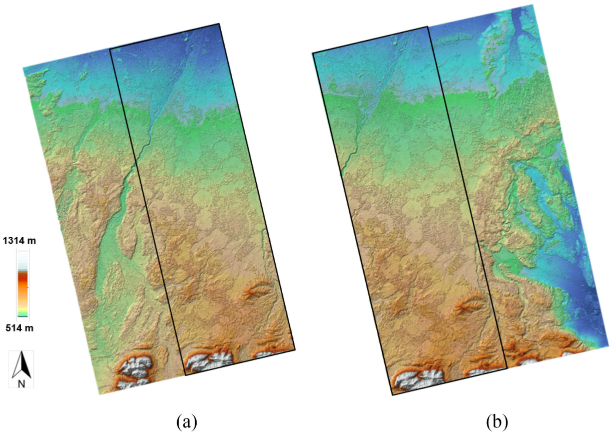


Fig. 2. Two nominal acquisitions of TanDEM-X raw DEMs over the Munich area. (a) 1145180 tile. (b) 1023491 tile.

± 20 cm. The final reference DEM is achieved by interpolation in a grid with pixel spacing as same as input TanDEM-X DEMs.

The results of raw DEM fusion using TV-L_1 and Huber models for study areas are presented in Table II. The regularization parameter is calculated using the L-curve method. For comparison with the common fusion method, the results of fusion by WA are also provided. The DEM quality after and before fusion was evaluated by statistical metrics, mean, root mean square error (RMSE), mean absolute error (MAE), normal median absolute deviation (NMAD), and standard deviation (STD).

In addition to the statistical analysis, to evaluate the performance of variational models, the residual maps of the input DEMs and the fused DEMs achieved by different methods for the industrial and inner city 1 study areas are displayed in Fig. 4. The results illustrate that using variational models in the fusion process can finally improve the quality of the TanDEM-X DEM over the quality achievable with classic WA. It is explicitly displayed on residual maps that variational modes can finally reduce the noise effects and also makes the footprints of buildings more apparent than WA. Furthermore, fusing ascending and descending DEMs can improve the DEM quality in particular for the shadow- and layover-affected areas in which significant errors occur.

B. Fusion of TanDEM-X Raw DEMs With Different HoAs

In the first experiment, the study areas were selected from two TanDEM-X raw DEMs that have nearly similar properties. Both DEMs were acquired in the same orbit, same look directions, and also with nearly the same incidence angles and HoAs.

As an additional experiment, we investigate the performance of variational models for fusing TanDEM-X raw DEMs with different HoAs over urban areas. For this purpose, one experimental ITP raw DEM with different HoA over Munich city in Germany is considered. It should be noted that this product has not been used in final global DEM generation but in this study is applied for implementing an experiment of fusing rawDEMs with different HoAs. The specifications of this raw DEM are provided in Table III. Fig. 5 also provides a depiction of the new raw DEM, which is acquired over the same location as tile 1023491 with identical overlap.

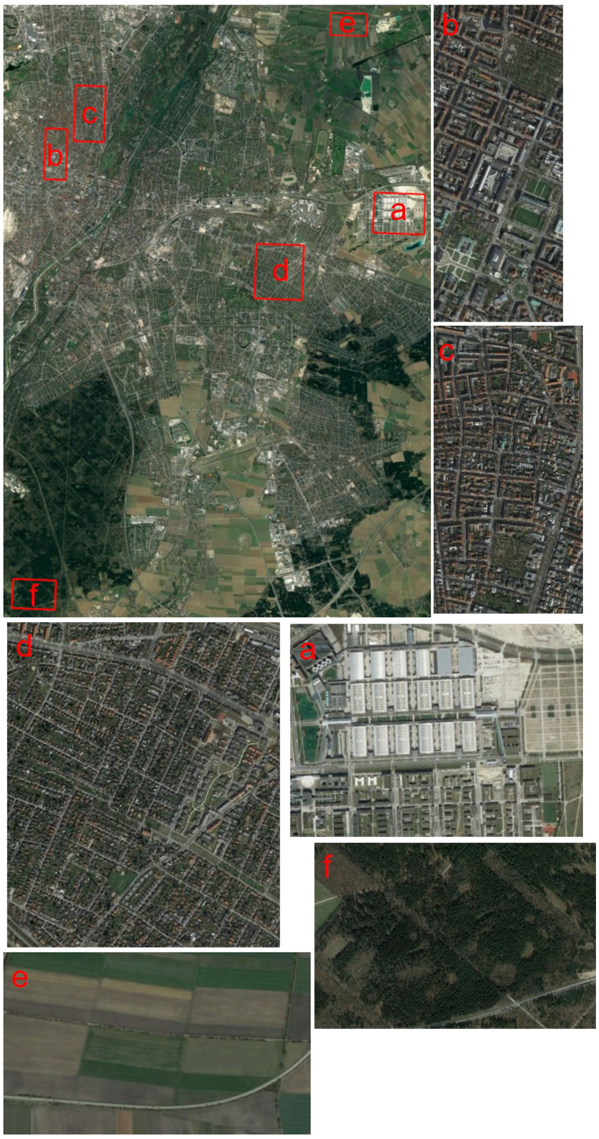


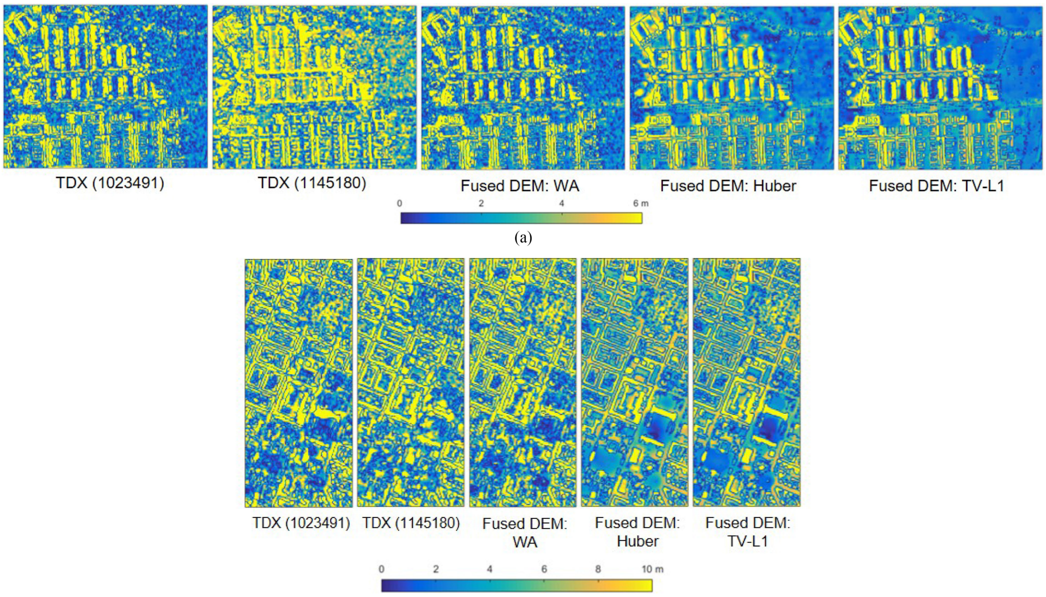
Fig. 3. Display of the study subsets selected from different urban land types for TanDEM-X raw DEM fusion over Munich city. a) Industrial area (1.5 km × 1.2 km). b) Inner city 1 (1.5 km × 0.6 km). c) Inner city 2 (1.6 km × 0.9 km). d) Residential area (1.6 km × 1.3 km). e) Agricultural (1.05 km × 0.6 km). f) Forested (1.25 km × 0.85 km).

The main property that discriminates this tile from those introduced in the previous section is its bigger HoA. Regarding nearly similar incidence angle and slant range, the larger value for HoA means this tile is derived from data takes that were acquired with a shorter baseline is considered helpful in areas where PU errors are dominant [34]. On the other hand, the quality and resolution of this DEM is lower than those with smaller

TABLE II
HEIGHT ACCURACY (IN METER) OF THE TANDEM-X DATA BEFORE AND AFTER DEM FUSION IN THE DIFFERENT STUDY AREAS OVER MUNICH

Study area	DEM		Mean	RMSE	MAE	NMAD	STD
Industrial	Raw DEM	id: 1023491	0.71	4.40	3.08	2.37	4.34
		id: 1145180	0.71	4.64	3.27	3.01	4.58
	Fused DEM	WA	0.77	4.16	2.93	2.24	4.09
		TV-L_1	**0.69**	**3.67**	**2.69**	**2.03**	**3.60**
		Huber	0.71	3.74	2.84	2.40	3.67
Inner 1	Raw DEM	id:1023491	0.78	7.79	5.95	6.49	7.75
		id:1145180	0.78	8.08	6.30	7.15	8.04
	Fused DEM	WA	0.84	7.51	5.83	6.49	7.46
		TV-L_1	**0.77**	**6.11**	**5.00**	5.72	**6.06**
		Huber	0.78	6.14	5.09	**5.67**	6.09
Inner 2	Raw DEM	id:1023491	0.18	7.00	5.44	6.36	7.00
		id:1145180	0.18	7.16	5.57	6.51	7.16
	Fused DEM	WA	0.20	6.82	5.33	6.23	6.82
		TV-L_1	**0.12**	5.83	**4.78**	**6.16**	5.83
		Huber	0.18	**5.82**	4.82	6.23	**5.82**
Residential	Raw DEM	id:1023491	0.95	2.68	2.10	2.05	2.50
		id:1145180	0.95	2.92	2.25	2.31	2.76
	Fused DEM	WA	0.96	2.61	2.05	1.99	2.43
		TV-L_1	**0.89**	**2.41**	**1.96**	**1.98**	**2.24**
		Huber	0.95	2.44	1.98	**1.98**	**2.24**
Agricultural	Raw DEM	id:1023491	0.13	0.86	0.57	0.59	0.84
		id:1145180	0.13	1.64	1.13	1.20	1.64
	Fused DEM	WA	0.14	0.78	0.51	0.54	0.76
		TV-L_1	**0.06**	**0.55**	**0.29**	**0.20**	**0.54**
		Huber	0.13	0.72	0.48	0.47	0.71
Forested	Raw DEM	id:1023491	**2.25**	4.84	3.54	3.46	4.28
		id:1145180	**2.25**	4.58	3.36	3.24	3.99
	Fused DEM	WA	2.28	4.51	3.30	3.17	3.89
		TV-L_1	**2.25**	**4.34**	**3.18**	**3.09**	**3.71**
		Huber	**2.25**	4.36	3.21	3.12	3.73

The bold values indicate the best results.



(a)

(b)

Fig. 4. Absolute residual maps of the initial input raw DEMs and the fused DEMs obtained by different approaches for the industrial areas. (a) Inner city areas. (b) Study areas over Munich.

TABLE III
PROPERTIES OF THE NONOFFICIAL TANDEM-X RAW DEM TILE FOR MUNICH AREA

TanDEM-X raws DEMs: Munich area	
Acquisition Id	1058842
Acquisition mode	Stripmap
Center incidence angle	38.33 °
Equator crossing direction	Ascending
Look direction	Right
Polarization	HH
Height of ambiguity	72.02
Pixel spacing	0.2 arcsec
HEM mean	2.58

HoAs. Comparing the raw tiles displayed in Figs. 2 and 5 confirms a drop of quality of DEM, i.e., with much more noise, with id 1058842, which has larger HoA.

In this experiment, a study subset is extracted from an area that has lots of inconsistent heights due to PU errors. For this aim, a relatively large subset from an urban area, which is covered by trees and also includes a river crossing, is selected. Fig. 6 displays the selected study area suffering from PU errors. The corresponding DEM data are derived from tiles 1023491 and 1058842 with HoAs about 45 m and 72 m, respectively.

The PU errors appearing in this subset originate from the volume decorrelation phenomenon that happens in an area covered by trees (like the selected study subset) and also a coherence change due to transition from dry land to water (river). PU errors typically are at the range of multiples of the HoA value. The inconsistent heights can be determined by [8]

$$dh_{\text{th}} = 0.75 \times \min(|HoA|) - 4. \tag{11}$$

Those height residuals bigger than dh_{th} are denoted as inconsistent height values emerging because of PU errors.

Table IV collects the results of fusing DEMs with different HoAs in the selected study area. Again, the accuracy was evaluated respective to a reference DSM interpolated from a point cloud with high density (more than eight points per square meters). Moreover, Table V compares the fused DEMs with different approaches and initial DEMs in terms of number of PU errors, maximum and minimum height residuals. The PU threshold for each DEM is computed based on the respective HoA value using (11). It is obvious that the DEM 1058842 has lower number of PU errors because of larger HoA, but for DEM fusion quality analysis, the minimum value of HoAs (here 45.81) is considered to enumerate the number of PU errors. It should be noted that mean values presented in Tables II and IV do not present the real level of canopy penetration of the *X*-band radar signal. In our previous study [35], we found some amount of vegetation penetration to remain after DEM coregistration.

The results from Tables IV and V demonstrate the efficiency of the Huber model for fusion of two tiles of TanDEM-X raw DEMs in the problematic area. The results show that using the Huber model can significantly improve the RMSE of fused DEMs by up to nearly 2 m while the DEM quality enhancement by means of WA is not remarkable. Apart from this, the Huber model is absolutely more powerful than WA to reduce the PU errors. The maximum and minimum discrepancies also confirm the better performance of the Huber model to deal with PU errors in comparison to the other method. TV-L_1 also can decrease the noise effect in the final fused DEM and reduce the number of PU errors but the improvement is not as large as for the Huber model.

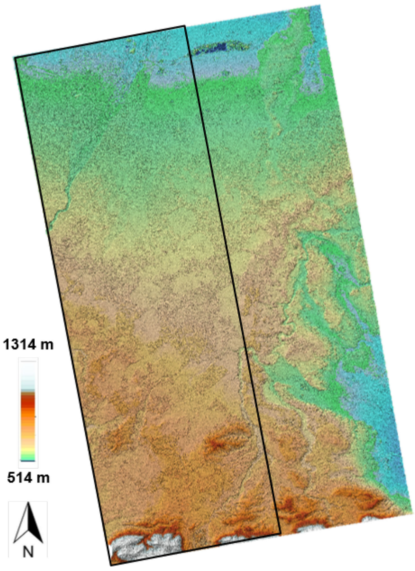


Fig. 5. Nonofficial TanDEM-X raw DEM tile produced with bigger HoA over the Munich area.

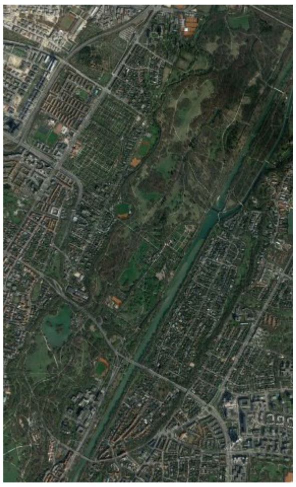


Fig. 6. Study subset selected for DEM fusion in a problematic area (4.5 km × 2.8 km).

C. Fusion of TanDEM-X Raw DEMs With Different Baseline Configuration

In the final experiment, we focus on the fusion of DEMs acquired by different baseline configurations including different orbit directions and HoAs. Table VI provides the properties of the tiles used for this experiment. The raw DEMs covering

TABLE IV
HEIGHT ACCURACY (IN METER) OF THE TANDEM-X DATA WITH DIFFERENT HOAS BEFORE AND AFTER DEM FUSION IN THE PROBLEMATIC STUDY AREA

DEM		Mean	RMSE	MAE	NMAD	STD
Raw DEM	id: 1023491	**-2.35**	10.77	8.46	10.10	10.51
	id: 1058842	**-2.35**	10.57	8.27	9.69	10.30
Fused DEM	WA	-2.37	10.45	8.23	9.81	10.17
	TV-L_1	-2.63	9.24	7.13	8.03	8.86
	Huber	**-2.35**	**8.60**	**6.70**	**7.65**	**8.277**

The bold values indicate the best results.

TABLE V
EFFECT OF DEM FUSION TO REDUCE THE NUMBER OF PU ERRORS USING TILES WITH DIFFERENT HOAS IN THE PROBLEMATIC STUDY AREA

DEM		HoA	PU Threshold	No. of PU Errors	Max Discrepancy	Min Discrepancy
Raw DEM	id: 1023491	45.81	30.36	2032	51.80	-73.13
	id: 1058842	72.02	50.01	51	58.82	-54.76
Fused DEM	WA	45.81	30.36	1339	50.74	-53.39
	TV-L_1	45.81	30.36	102	19.16	-33.76
	Huber	**45.81**	**30.36**	**0**	**16.97**	**-28.71**

The bold values indicate the best results.

TABLE VI
PROPERTIES OF THE NOMINAL ASCENDING AND DESCENDING TANDEM-X RAW DEM TILES OVER TERRASSA AND VACARISSES CITIES

	TanDEM-X raws DEMs	
Acquisition Id	1058683	1171358
Acquisition mode	Stripmap	Stripmap
Center incidence angle	33.71°	34.82°
Equator crossing direction	Ascending	Descending
Look direction	Right	Right
Polarization	HH	HH
Height of ambiguity	60.18 m	48.58 m
Pixel spacing	0.2 arcsec	0.2 arcsec
HEM mean	1.17 m	1.40

Terrassa and Vacarisses cities located in Spain were produced by ascending and descending acquisitions. In addition to orbit directions, the HoAs of tiles are also not similar to each other. Fig. 7 shows the TanDEM-X raw DEMs used in this study, which mostly covers difficult terrain, the common area is specified by black polygons. Due to morphologically difficult type of terrain, the acquisitions from ascending and descending flight paths have been applied for global DEM generation in this area. However, the study cities are located in the relatively flat part of the area common between tiles. Again, from these tiles, study subsets located in different land types were selected. Fig. 8 display each study subsets from different types extracted from the common area of ascending and descending raw tiles.

The results of fusing ascending and descending raw DEMs in different land types over urban area are provided in Table VII. The accuracy evaluation is performed by comparing each DEM respective to a LiDAR DSM, which was achieved by interpolation of the LiDAR point cloud presented by the ISPRS foundation as a benchmark [36]. On an average, the density of the point cloud is about one point per square meter.

The results of DEM fusion again illustrate that using variational models can increase the accuracy of the initial input raw DEMs. In urban study subsets, the performance of the Huber model is slightly better than TV-L_1 according to the statistical metrics, but their differences are not really significant. It can be concluded that both models produce similar results in terms of statistical measurements. In comparison to WA, variational models also give a more accurate DEM in urban areas and the agricultural subsets.

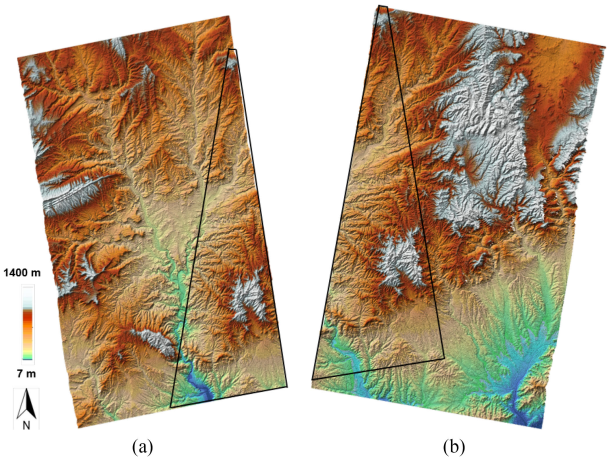


Fig. 7. (a) Ascending and (b) descending tiles of TanDEM-X raw DEMs produced over Terrassa and Vacarisses cities.

IV. DISCUSSION

In this study, the TV-L_1 and Huber variational models were implemented to fuse TanDEM-X raw DEMs over urban areas as well as surroundings. In particular, we investigated these models with respect to the fusion of raw DEMs produced from data takes with different baseline configurations and HoAs. In conclusion, the results demonstrated the efficiency of variational models in comparison to simple WA for the TanDEM-X raw DEM fusion. To clarify the role of smoothness constraint and data term, we

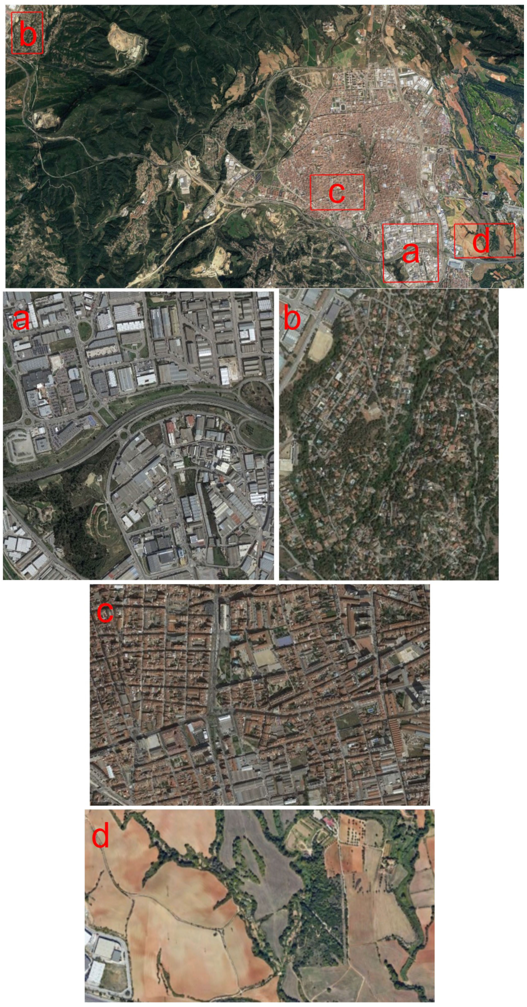


Fig. 8. Display of study subsets selected from different land types for raw TanDEM-X fusion over urban areas. a) Industrial area located in Terrassa (1.5 km × 1.4 km). b) Residential area located in Vacarisses (1.3 km × 0.9 km). c) Inner city subset located in Terrassa (1 km × 0.8 km). d) Agricultural area (1.5 km × 0.8 km).

carried out an experiment regarding the DEM quality improvement to be achieved by just carrying out TV-$L1$ denoising of a single input DEM. Comparing these results with those achieved by TV-$L1$ DEM fusion (which employs elevation data of at least two DEM tiles) revealed that fusion is always favorable (cf. Table VIII). Furthermore, it can be seen that in the industrial subset, the main improvement arises from the smoothness term, which is caused by the regular scene structure. However, adding another tile in a fusion manner can still improve the quality of the final DEM. In contrast, for the agricultural subset, the TV-$L1$ denoising could not change the DEM quality, while DEM fusion could finally produce a DEM with higher accuracy. Using more DEM tiles is furthermore vital for areas suffering from layover and shadowing effects or containing PU errors. More examples of these areas and requirement for employing several tiles can be found in [8]. In addition, regarding the strict quality control policy of the TanDEM-X mission, obtaining lower than 2 m relative height accuracy for slopes lower than 20% and better than 4 m for steeper slopes in each pixel means that the pixelwise TanDEM-X target accuracy can only be realized by DEM fusion. Table VIII provides some statistics relevant to TanDEM-X quality control indicating percentages of pixels with an accuracy better than 2 and 4 m as well as the percentage of pixels with accuracy worse than 4 m. The results confirm that DEM fusion can lead to obtaining more reliable pixels in comparison to just the denoising of single DEMs. Another important problem with using a single tile is the selection of the accurate subsets in different land types. As an example, in experiment 1, the accuracy of DEM with id 1145180 is higher than the accuracy of DEM 1023491 for the forested area while for other study subsets the quality of DEM 1023491 is better than for the other DEMs. As a result, in practice, it is beneficial to carry out DEM fusion in general, as this always improves the quality of the final DEM.

A. *Use of TV-Based Variational Models*

The main property of TV-based models is to reduce the effect of noise by minimizing the TV term. It should be noted that both data and regularization terms in the energy functional defined for TV-L_1 and Huber models are positive terms. Choosing TV as a regularization term leads to preserving the beneficial high-frequency image contents such as footprints of buildings while minimizing its value through the fusion causes to reduce the effects of undesirable noise. Fig. 9 shows the performance of TV-based variational models in comparison to WA in a 3-D view. The displayed patch was selected from an industrial area located in Munich, which was used in the first experiment.

The 3-D display of the fused DEMs clearly shows that the TV-based model can reduce the noise effect and excellently reveal the edges while the WA-based fused DEM still suffers from noise effects. As displayed, the Huber model produces a smoother output in comparison to TV-L_1 because of mixing the quadratic norm and the L_1 norm to form data and regularization terms. Since the quadratic norm tends to penalize the high-frequency contents more severe than L_1, it leads to DEMs with more smoother edges. Apart from the type of norm used to form an energy functional, the amount of smoothing induced by TV-based variational models depends on the regularization parameter, which trades off between the TV term as a regularization term and the data fidelity term. While only one regularization parameter is required to be tuned for DEM fusion by TV-L_1, using the Huber model for fusion demands to tune three parameters. Selecting different thresholds to form the norms used in the Huber model changes the amount of smoothness that emerges in the final output of DEM fusion. Fig. 10 displays the effect of changing one of the parameters while the others are constant on the final output. Selecting small α, which is used for a data term, does not severely penalize discrepancies between the initial DEMs and the desired output strongly, i.e., giving an output fused DEM with more similarity to input data. In contrast, increasing α penalizes the discrepancies intensively and the optimization process tries to lower the total energy that provides a smoother DEM at the end. An identical interpretation can be derived for β while this parameter performs in a

TABLE VII
HEIGHT ACCURACY (IN METER) OF THE ASCENDING AND DESCENDING TANDEM-X DATA BEFORE AND AFTER DEM FUSION IN THE DIFFERENT STUDY AREAS OVER VACARISSES AND TERRASSA

Study area	DEM		Mean	RMSE	MAE	NMAD	STD
Industrial	Raw DEM	1058683	-0.19	3.49	2.49	2.60	3.48
		1171358	-0.19	3.56	2.44	2.34	3.55
	Fused DEM	WA	-0.26	3.06	2.13	2.07	3.05
		TV-L_1	-0.34	2.92	**2.09**	**2.07**	2.90
		Huber	**-0.19**	**2.89**	2.10	2.14	**2.88**
Inner	Raw DEM	1058683	-0.78	5.05	3.52	3.70	4.99
		1171358	-0.78	5.11	3.53	3.62	5.05
	Fused DEM	WA	-0.76	4.66	3.22	**3.36**	4.59
		TV-L_1	-0.91	4.35	**3.08**	3.40	**4.25**
		Huber	**-0.78**	**4.34**	3.13	3.52	4.27
Residential	Raw DEM	1058683	-0.54	4.24	3.11	3.19	4.20
		1171358	-0.54	4.42	3.21	3.26	4.38
	Fused DEM	WA	-0.62	3.94	2.87	2.83	3.90
		TV-L_1	-0.76	3.96	2.88	2.77	3.88
		Huber	**-0.54**	**3.86**	**2.86**	**2.74**	**3.82**
Agricultural	Raw DEM	1058683	0.44	2.38	1.68	1.71	2.34
		1171358	0.44	1.93	1.23	0.98	1.88
	Fused DEM	WA	0.35	**1.60**	**1.04**	0.83	1.57
		TV-L_1	**0.27**	**1.60**	**1.04**	**0.78**	1.59
		Huber	0.44	1.62	1.12	0.91	**1.56**

The bold values indicate the best results.

TABLE VIII
COMPARISON OF TV-L_1 DENOISING AND TV-L_1 DEM FUSION IN INDUSTRIAL AND AGRICULTURAL AREAS USED IN THE FIRST EXPERIMENT

Study area	Strategy	RMSE	MAE	NMAD
Residential	TV-L_1 DEM denoising	3.88	2.88	2.20
	TV-L_1 DEM fusion	**3.67**	**2.69**	**2.03**
		Error $<$ 2m	Error $<$ 4m	Error $>=$ 4m
Residential	TV-L_1 DEM denoising	49 %	78 %	22 %
	TV-L_1 DEM fusion	**54 %**	**81 %**	**19%**
Study area	Strategy	RMSE	MAE	NMAD
Agricultural	TV-L_1 DEM denoising	0.86	0.57	0.59
	TV-L_1 DEM fusion	**0.55**	**0.29**	**0.20**
		Error $<$ 2m	Error $<$ 4m	Error $>=$ 4m
Agricultural	TV-L_1 DEM denoising	95 %	100 %	0 %
	TV-L_1 DEM fusion	**100 %**	**100 %**	**0%**

The bold values indicate the best results.

reverse manner because it is used to form the regularization term. It should be noted that the regularization parameter γ trades off between two terms in functional energy that means by increasing γ, the effect of TV will become lower such that ultimately smoother DEM is produced. Appropriately tuning the regularization parameter and the Huber model thresholds also influences the accuracy of the final fused DEM. The effect of Huber model parameter values on the final accuracy of DEM fusion is depicted in Fig. 11. Different methods can be used for tuning the regularization parameter. One option is to learn it from data if some training data are available. Another option is to use the L-curve approach [32]. Finally, the parameter can be manually selected based on a visual analysis of different output DEMs.

B. Fusion Over Different Land Types

In this study, the TanDEM-X DEM fusion by variational models was implemented over different land types that are typically found in urban areas and in their surroundings. Fig. 12 depicts the accuracy improvement (in meters) by means of different fusion algorithm respective to the quality of the initial DEMs for each study land type used in the first experiment (see Section III-A). Similarly, Fig. 13 compares the performance of fusion methods for different land types that were used in the third experiment (see Section III-C). It should be noted that since both variational model have similar performance in terms of RMSE, for each plot in Figs. 12 and 13, just the performance of the best variational model is compared to WA.

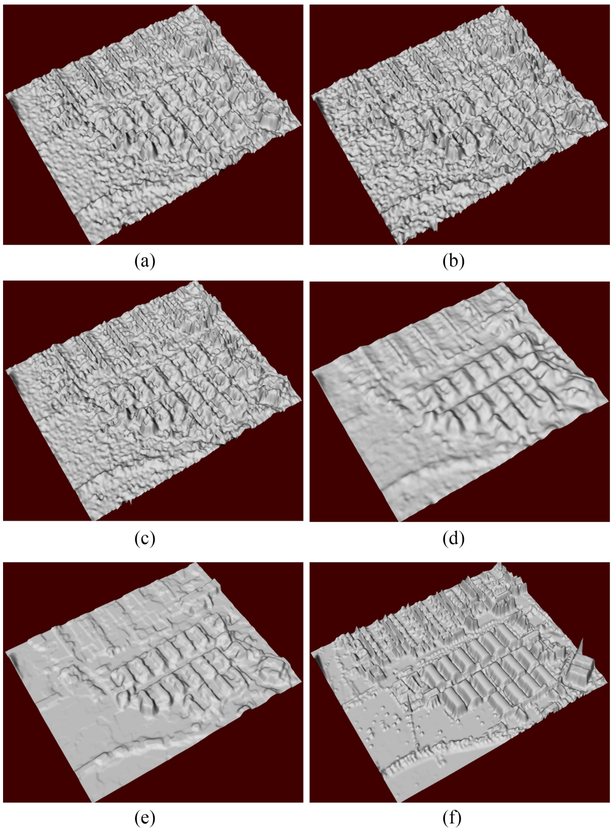


Fig. 9. 3-D display of initial TanDEM-X raw data and the results of DEM fusions using different methods in the industrial area used in the first experiment. (a) TanDEM-X (tile a). (b) TanDEM-X (tile b). (c) WA. (d) Huber model. (e) TV-L1 model. (f) LiDAR.

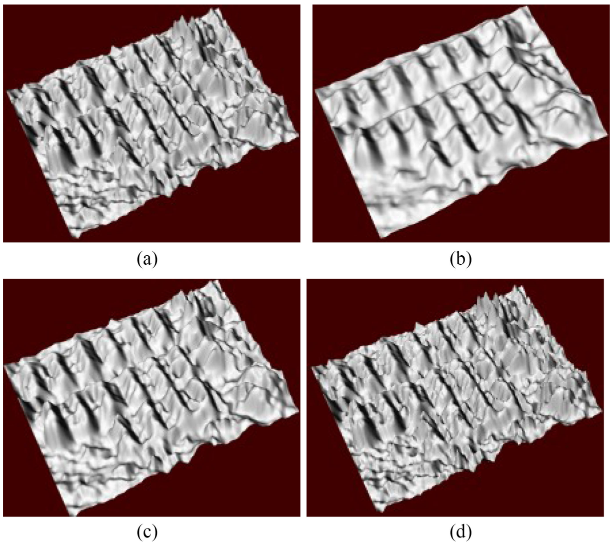


Fig. 10. Effect of varying Huber models' parameters on the final DEM. (a) $\gamma = 1, \alpha = 0.5, \beta = 1$. (b) $\gamma = 1, \alpha = 10, \beta = 1$. (c) $\gamma = 1, \alpha = 1, \beta = 0.5$. (d) $\gamma = 1, \alpha = 1, \beta = 10$.

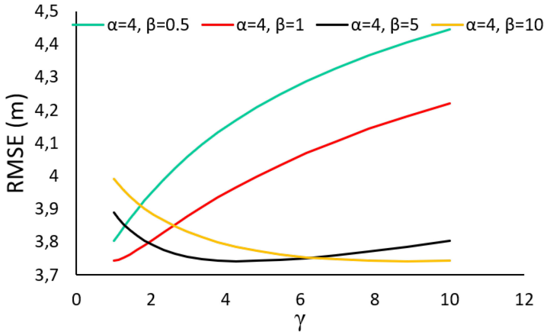


Fig. 11. Influence of different Huber norm parameters on the RMSE of fused DEM.

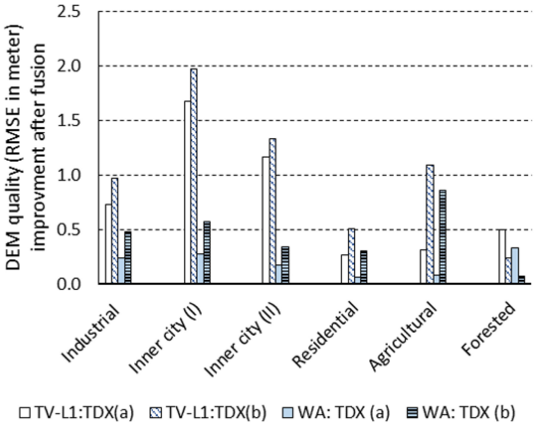


Fig. 12. Improvement of the TanDEM-X DEM tiles (a), (b) using variational models (here, TV-L1) in comparison to WA in different study areas located in Munich. The bars indicate the difference between the RMSE of input TanDEM-X DEM and final fused DEM. (a) Refers to tile 1023491. (b) Refers to tile 1145180.

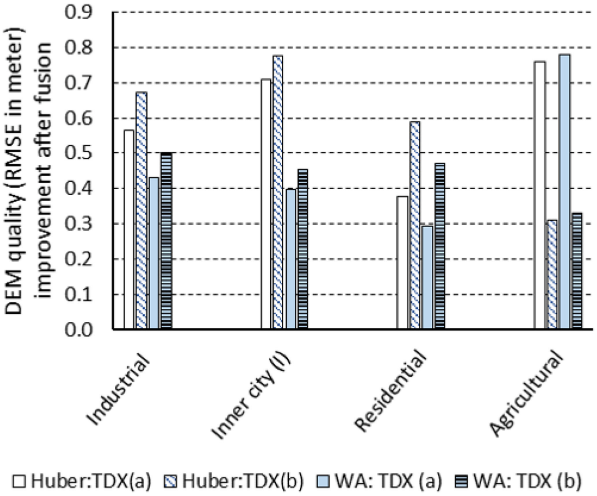


Fig. 13. Improvement of the TanDEM-X DEM tiles (a), (b) using variational models (here, Huber) in comparison to WA in different study areas located in Vacarisses and Terrassa. The bars indicate the difference between the RMSE of input TanDEM-X DEM and final fused DEM. (a) Refers to tile 1058683. (b) Refers to tile 1171358.

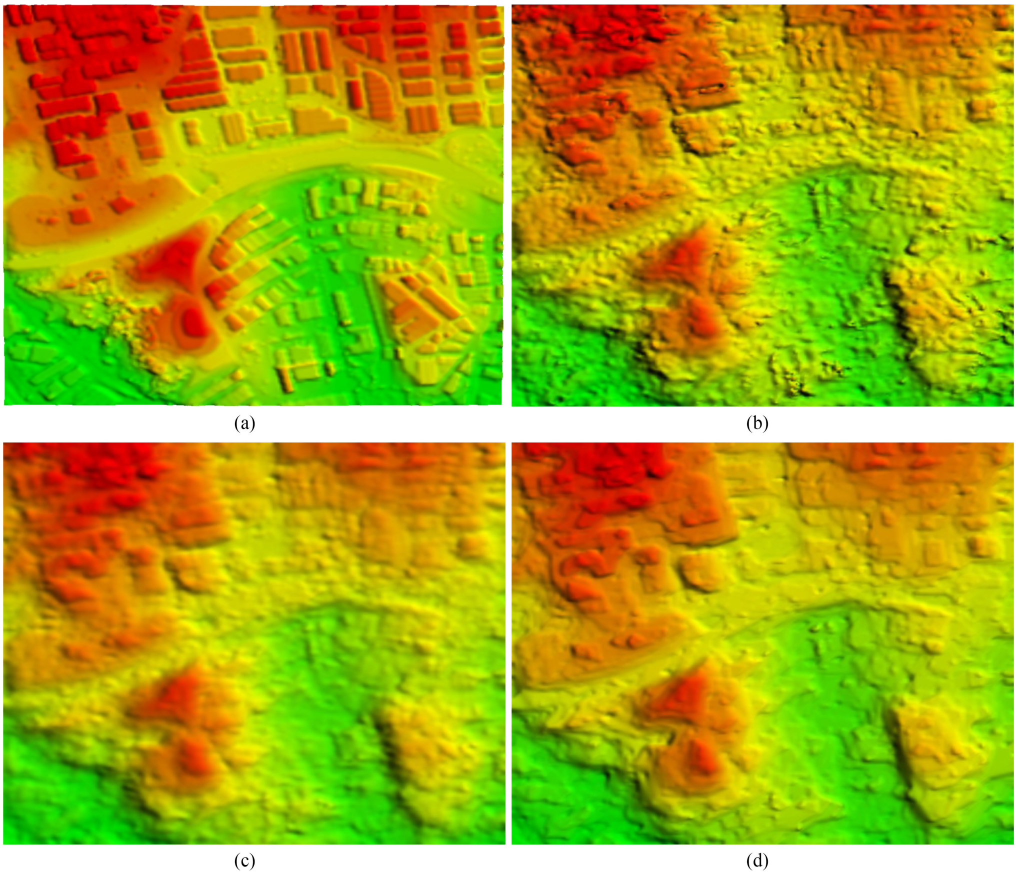


Fig. 14. DEMs produced by fusing ascending and descending DEMs over industrial study area located in Terrassa. (a) LiDAR. (b) WA. (c) Huber. (d) TV-L_1.

The plots demonstrate that variational models exhibit maximum efficiency in inner city land types in both experiments while WA has a nearly similar performance in different land types. The lowest accuracy improvement by variational models is for residential subset and nonurban study areas. The inner city land type includes a lot of building footprints that mostly appear as noisy edges because of inherent properties of SAR sensor imaging. Consequently, using the TV-based variational model can significantly improve the DEM quality in these areas. On the other hand, residential subset areas include single family, small homes usually located in a sparse pattern, and the footprint of buildings, which cannot appear as a strong edge in TanDEM-X raw DEM due to resolution restriction of data takes acquired in the stripmap mode. In nonurban areas, the edginess is usually lower than in urban subsets. Thus, the smoothness term of the variational models has lower performance in those kinds of land types. However, the quality of the final DEM still increases due to the DEM fusion encoded in the data term.

C. *Effect of Geometry*

While most urban areas covered by global TanDEM-X dataset are generated by two nominal acquisitions that mostly have similar HoAs and geometries, we also investigated the fusion of several TanDEM-X DEMs with different properties to investigate the performance of variational models for these data. In the first experiment, the results identified that the variational models can perfectly fuse the raw DEMs with nearly similar baseline configuration and HoAs acquired over urban areas. The output is a DEM with higher accuracy and more enhanced building footprints. However, the Huber model generates a smoother DEM at the end.

A significant result was yielded for problematic areas where the effects of PU errors are dominant. The selected study subset (see Fig. 6) is mostly affected by noise because of the volume decorrelation due to trees and the low coherence due to river. For these problematic areas fusing one DEM with nominal HoA to another DEM with larger HoA is more useful to reduce the effect of PU errors. In this experiment, fusing two DEMs with different HoAs by using the Huber model could substitute inconsistent heights with logical values and also resulted in a more accurate DEM. This proves, in addition to the DEM fusion methodology, that selecting appropriate raw DEM tiles dependent to problem is significant for a successful fusion. Among variational models, TV-L_1 can decrease the number of PU errors and improve the accuracy but more quality enhancement is achieved by the Huber model. Since, the Huber model also uses the quadratic norm, it produces a smoother fused DEM while TV-L_1 tends to save more high-frequency contents that can also be caused by noise.

Fusing ascending and descending DEMs in problematic areas reduces the layover and shadow effects in the final fused DEM. Consequently, in the final experiment, two ascending and descending DEMs with different HoAs were fused. As shown in plots 12 and 13, in comparison to results of fusing DEMs with similar baseline configuration and HoAs, the variational models lead to least significant quality improvement in the final fused DEM in terms of RMSE. However, a display of an exemplary study subset (industrial area) in Fig. 14 demonstrates the efficiency of variational models in comparison to WA for fusing these types of DEMs. For making a correct judgment about the performance of variational models on fusing ascending and descending DEMs versus DEMs with similar flight paths, two DEMs with similar baseline configuration and HoA from the study areas are required. Theoretically, apart from the DEM fusion method, using ascending and descending DEMs instead of using DEMs with similar orbit directions improves the final DEM quality in the difficult terrains and problematic areas such as urban areas that are under the shadow and layover effects. In practice, it is confirmed in [8] that using ascending and descending raw TanDEM-X DEMs can produce highly accurate fused DEM at the end in the shadow- and layover-affected areas.

V. Conclusion

In this paper, we proposed to apply TV-based variational models (TV-L_1 and Huber models) for TanDEM-X raw DEM fusion at the phase of DEM mosaicking instead of WA. The main focus of this study was to enhance final DEMs in urban areas where the footprints of buildings are influenced by noise effects due to SAR imaging properties. For this purpose, different study subsets were selected from different land types, which mostly are explored over urban areas and surroundings. Apart from this, DEM fusion was investigated for raw DEMs with different geometries. At first, two nominal acquisitions with similar baseline configurations and HoAs were fused over different land types. In the next experiment, two raw DEMs with different HoAs were fused over a problematic terrain that suffers from PU errors. At the end, two DEMs with ascending and descending orbit directions as well as with different HoAs were used. In all experiments, it was demonstrated that using variational models leads to DEMs with higher quality. A great performance of the Huber model was recorded for fusing two raw DEMs with different HoAs over the selected problematic area. Also, in urban areas, variational models with reducing the noise effect and enhancing the outlines of buildings, absolutely performs better than WA. However, the Huber model tends to provide a smoother fused DEM than TV-L_1. The results also demonstrated that the variational models, particularly TV-L_1, could improve the quality of DEMs significantly in comparison to WA. Using variational models could improve the DEM quality by up to 2 m particularly in inner city subsets. In conclusion, carrying out TanDEM-X raw DEM fusion using variational models with an ability to enhance the building footprints and other useful high-frequency contents along with smoothing the noise, finally produced a DEM with higher quality.

Acknowledgment

The authors would like to thank Dr. Fritz of DLR and Dr. Baier for providing the TanDEM-X raw DEM for the Munich area; and the Bavarian Surveying Administration for providing the LiDAR data of Munich.

References

[1] E. Rodriguez, C. S. Morris, and J. E. Belz, "A global assessment of the SRTM performance," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 3, pp. 249–260, 2006.

[2] G. Krieger *et al.*, "TanDEM-X: A satellite formation for high-resolution SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3317–3341, Nov. 2007.

[3] P. Rizzoli *et al.*, "Generation and performance assessment of the global TanDEM-X digital elevation model," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, no. Suppl C, pp. 119–139, 2017.

[4] T. Fritz, H. Breit, C. Rossi, U. Balss, M. Lachaise, and S. Duque, "Interferometric processing and products of the TanDEM-X mission," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2010, pp. 1904–1907.

[5] J. H. Gonzlez *et al.*, "Bistatic system and baseline calibration in TanDEM-X to ensure the global digital elevation model quality," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 3–11, 2012.

[6] A. Gruber, B. Wessel, M. Huber, and A. Roth, "Operational TanDEM-X DEM calibration and first validation results," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 39–49, 2012.

[7] B. Wessel *et al.*, "Design of the DEM mosaicking and calibration processor for TanDEM-X," in *Proc. 7th Euro. Conf. Synthetic Aperture RADAR*, Jun. 2008, pp. 1–4.

[8] A. Gruber, B. Wessel, M. Martone, and A. Roth, "The TanDEM-X DEM mosaicking: Fusion of multiple acquisitions using InSAR quality parameters," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 3, pp. 1047–1057, Mar. 2016.

[9] M. Schmitt and X. X. Zhu, "Data fusion and remote sensing: An ever-growing relationship," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 4, pp. 6–23, Dec. 2016.

[10] R. Deo, C. Rossi, M. Eineder, T. Fritz, and Y. S. Rao, "Framework for fusion of ascending and descending pass TanDEM-X raw DEMs," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 7, pp. 3347–3355, Jul. 2015.

[11] P. Reinartz, R. Müller, D. Hoja, M. Lehner, and M. Schroeder, "Comparison and fusion of DEM derived from SPOT-5 hrs and SRTM data and estimation of forest heights," in *Proc. Earsel Workshop 3D-Remote Sens. Earsel Symp.*, Porto, Portugal, Jun. 2005, p. (on CD ROM).

[12] D. Hoja and P. d'Angelo, "Analysis of DEM combination methods using high resolution optical stereo imagery and interferometric SAR data," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 37, 2009, p. (on CD ROM).

[13] A. Roth, W. Knopfle, G. Strunz, M. Lehner, and P. Reinartz, "Towards a global elevation product: combination of multi-source digital elevation models," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 34, no. 4, pp. 675–679, 2002.

[14] W. Knöpfle, G. Strunz, and A. Roth, "Mosaicking of digital elevation models derived by SAR interferometry," *Int. Arch. Photogramm. Remote Sens.*, vol. 32, pp. 306–313, 1998.

[15] D. Just and R. Bamler, "Phase statistics of interferograms with applications to synthetic aperture RADAR," *Appl. Opt.*, vol. 33, pp. 4361–4368, Jul. 1994.

[16] H. Bagheri, M. Schmitt, and X. X. Zhu, "Fusion of TanDEM-X and Cartosat-1 elevation data supported by neural network-predicted weight maps," *ISPRS J. Photogramm. Remote Sens.*, vol. 144, pp. 285–297, 2018.

[17] H. Papasaika, E. Kokiopoulou, E. Baltsavias, K. Schindler, and D. Kressner, "Fusion of digital elevation models using sparse representations," in *Proc. ISPRS Conf. Photogramm. Image Anal.*, 2011, pp. 171–184.

[18] C. Zach, T. Pock, and H. Bischof, "A globally optimal algorithm for robust TV-L1 range image integration," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[19] T. Pock, L. Zebedin, and H. Bischof, *TGV-Fusion*. Berlin, Germany: Springer, 2011, pp. 245–258.

[20] G. Kuschk, P. d'Angelo, D. Gaudrie, P. Reinartz, and D. Cremers, "Spatially regularized fusion of multiresolution digital surface models," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1477–1488, Mar. 2017.

[21] H. Bagheri, M. Schmitt, and X. X. Zhu, "Fusion of TanDEM-X and Cartosat-1 DEMs using TV-norm regularization and ANN-predicted weights," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2017, pp. 3369–3372.

[22] A. Gruen and D. Akca, "Least squares 3D surface and curve matching," *ISPRS J. Photogramm. Remote Sens.*, vol. 59, no. 3, pp. 151–174, 2005.

[23] M. Ravanbakhsh and C. S. Fraser, "A comparative study of DEM registration approaches," *J. Spatial Sci.*, vol. 58, no. 1, pp. 79–89, 2013.

[24] M. Rumpler, A. Wendel, and H. Bischof, "Probabilistic range image integration for DSM and true-orthophoto generation," in *Proc. Scand. Conf. Image Anal.*, 2013, pp. 533–544.

[25] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D, Nonlinear Phenomena*, vol. 60, pp. 259–268, Nov. 1992.

[26] M. Nikolova, "A variational approach to remove outliers and impulse noise," *J. Math. Imag. Vis.*, vol. 20, pp. 99–120, Jan. 2004.

[27] A. N. Tikhonov, "On the stability of inverse problems," in *Proc. Dokl. Akad. Nauk SSSR*, 1943, vol. 39, pp. 195–198.

[28] T. F. Chan and S. Esedoglu, "Aspects of total variation regularized L1 function approximation," *SIAM J. Appl. Math.*, vol. 65, no. 5, pp. 1817–1837, 2005.

[29] T. Chan, S. Esedoglu, F. Park, and A. Yip, *Total Variation Image Restoration: Overview and Recent Developments*. Boston, MA, USA: Springer, 2006.

[30] P. J. Huber, *Robust Statistics*. Berlin, Germany: Springer, 2011, pp. 1248–1251.

[31] R. Perko and C. Zach, "Globally optimal robust DSM fusion," *Euro. J. Remote Sens.*, vol. 49, no. 1, pp. 489–511, 2016.

[32] P. C. Hansen and D. P. OLeary, "The use of the l-curve in the regularization of discrete ill-posed problems," *SIAM J. Scientific Comput.*, vol. 14, no. 6, pp. 1487–1503, 1993.

[33] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, pp. 120–145, May 2011.

[34] M. Martone, B. Bräutigam, P. Rizzoli, C. Gonzalez, M. Bachmann, and G. Krieger, "Coherence evaluation of TanDEM-X interferometric data," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 21–29, 2012.

[35] H. Bagheri, M. Schmitt, and X. X. Zhu, "Uncertainty assessment and weight map generation for efficient fusion of TanDEM-X and Cartosat-1 DEMS," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLII-1/W1, pp. 433–439, 2017.

[36] ISPRS. Satellite stereo benchmark. 2018. [Online]. Available: ftp://ftp.dlr.de/download/evaluation/index benchmark.htm

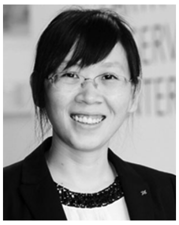
Hossein Bagheri received the master's degree in photogrammetry and remote sensing engineering from Tafresh University, Tafresh, Iran, in 2013. He is currently working toward the Ph.D. degree with the chair of Signal Processing in Earth Observation from Technical University of Munich, Munich, Germany.

In 2015, he spent six months with the Photogrammetry Group, Department of Surveying and Geomatics Engineering, University of Tehran, Tehran, Iran. His research interests include multisensor data fusion applied to SAR and optical data, 3-D reconstruction, and digital elevation models.

Michael Schmitt (S'08–M'14–SM'16) received the Dipl.-Ing. degree in geodesy and geoinformation and the Dr.-Ing. degree in remote sensing from the Technical University of Munich (TUM), Munich, Germany, in 2009 and 2014, respectively.

Since 2015, he has been a Senior Researcher and the Deputy Head with the Professorship for signal processing, Earth Observation, TUM. In 2016, he was a Guest Scientist with the University of Massachusetts Amherst, Amherst, MA, USA. His research interests include signal and image processing as well as machine learning for the extraction of information from remote sensing data, data fusion with emphasis on the joint exploitation of optical and radar data, 3-D reconstruction by techniques, such as SAR interferometry, SAR tomography, radargrammetry, or photogrammetry, and millimeter-wave SAR remote sensing.

Dr. Schmitt is a Co-Chair of the International Society for Photogrammetry and Remote Sensing Working Group I/3 on SAR and Microwave Sensing. He frequently serves as a Reviewer for a number of renowned international journals. In 2013 and 2015, he was elected as the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS Best Reviewer, leading to his appointment as an Associate Editor of the journal in 2016.



Xiao Xiang Zhu (S'10–M'12–SM'14) received the master's (M.Sc.) degree, her Doctor of Engineering (Dr.-Ing.) degree, and her Habilitation in the field of signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She is currently the Professor for signal processing in earth observation (www.sipeo.bgu.tum.de) with the Technical University of Munich (TUM), Munich, Germany and German Aerospace Center (DLR), Cologne, Germany; the Head of the Department "EO Data Science" at DLR's Earth Observation Center; and the Head of the Helmholtz Young Investigator Group "SiPEO" at DLR and TUM. She was a Guest Scientist or Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, the University of Tokyo, Tokyo, Japan, and the University of California, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. Her main research interests are remote sensing and earth observation, signal processing, machine learning, and data science, with a special application focus on global urban mapping.

Dr. Zhu is a member of young academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

A.3 H. Bagheri, M. Schmitt, P. d’Angelo, and X. X. Zhu. A framework for SAR-optical stereogrammetry over urban areas. In: ISPRS Journal of Photogrammetry and Remote Sensing 146 (2018), pp. 389–408.

ISPRS Journal of Photogrammetry and Remote Sensing 146 (2018) 389–408

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A framework for SAR-optical stereogrammetry over urban areas

Hossein Bagheri[a], Michael Schmitt[a], Pablo d'Angelo[b], Xiao Xiang Zhu[a,b,*]

[a] Signal Processing in Earth Observation, Technical University of Munich, Munich, Germany
[b] Remote Sensing Technology Institute, German Aerospace Center, Oberpfaffenhofen, Wessling, Germany

ARTICLE INFO

Keywords:
Epipolarity constraint
Multi-sensor block adjustment
Dense image matching
3D reconstruction
SAR-optical stereogrammetry

ABSTRACT

Currently, numerous remote sensing satellites provide a huge volume of diverse earth observation data. As these data show different features regarding resolution, accuracy, coverage, and spectral imaging ability, fusion techniques are required to integrate the different properties of each sensor and produce useful information. For example, synthetic aperture radar (SAR) data can be fused with optical imagery to produce 3D information using stereogrammetric methods. The main focus of this study is to investigate the possibility of applying a stereogrammetry pipeline to very-high-resolution (VHR) SAR-optical image pairs. For this purpose, the applicability of semi-global matching is investigated in this unconventional multi-sensor setting. To support the image matching by reducing the search space and accelerating the identification of correct, reliable matches, the possibility of establishing an epipolarity constraint for VHR SAR-optical image pairs is investigated as well. In addition, it is shown that the absolute geolocation accuracy of VHR optical imagery with respect to VHR SAR imagery such as provided by TerraSAR-X can be improved by a multi-sensor block adjustment formulation based on rational polynomial coefficients. Finally, the feasibility of generating point clouds with a median accuracy of about 2 m is demonstrated and confirms the potential of 3D reconstruction from SAR-optical image pairs over urban areas.

1. Introduction

Three-dimensional reconstruction from remote sensing data has a range of applications across different fields, such as urban 3D modeling and management, environmental studies, and geographic information systems. Manifold high-resolution sensors in space provide the possibility of reconstructing natural and man-made landscapes over large-scale areas. Conventionally, 3D reconstruction in remote sensing is either based on exploiting phase information provided by interferometric SAR, or on space intersection in the frame of photogrammetry with optical images or radargrammetry with SAR image pairs. In all these stereogrammetric approaches, at least two overlapping images are required to extract 3D spatial information. Both photogrammetry and radargrammetry, however, suffer from several drawbacks. Photogrammetry using high-resolution optical imagery is limited by relatively poor absolute localization accuracy and cloud effects, whereas radargrammetry suffers from the difficulty of image matching for severely different oblique viewing angles.

On the other hand, the huge archives of high-resolution SAR images provided by satellites such as TerraSAR-X and the regular availability of new data alongside archives of high-resolution optical imagery provided by sensors such as WorldView provide a great opportunity to investigate data fusion pipelines for producing 3D spatial information (Schmitt and Zhu, 2016). As relatively few studies have dealt with 3D reconstruction from SAR-optical image pairs (Bloom et al., 1988; Raggam et al., 1994; Wegner et al., 2014), there has been no investigation into the feasibility of a dense multi-sensor stereo pipeline as known from photogrammetric computer vision yet. This paper investigates the possibility of implementing such a pipeline, and describes all processing steps required for 3D reconstruction from very-high-resolution (VHR) SAR-optical image pairs.

In detail, this paper discusses both an epipolarity constraint and a bundle adjustment formulation for SAR-optical multi-sensor stereogrammetry first. Regarding the complicated radiometric relationship between SAR and optical imagery, the epipolarity constraint accelerates the matching process and helps to identify reliable and correct conjugate points (Morgan et al., 2004; Scharstein et al., 2001). For this objective, we first demonstrate the existence of an epipolarity constraint for SAR-optical imagery by reconstructing the rigorous geometry models of SAR and optical sensors using both collinearity and range-Doppler relationships. We prove that a SAR-optical epipolarity constraint can be rigorously modeled using the sensor geometries. Subsequently, rational polynomial coefficients (RPCs) are fitted to the SAR sensor geometry to ease further processing steps. Consequently,

[*] Corresponding author.
E-mail address: xiaoxiang.zhu@dlr.de (X.X. Zhu).

https://doi.org/10.1016/j.isprsjprs.2018.10.003
Received 20 July 2018; Received in revised form 28 September 2018; Accepted 9 October 2018
0924-2716/ © 2018 The Authors. Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

epipolar curves can be established using projection and back-projection from SAR imagery to terrain and then from terrain to optical imagery using RPCs. In addition, the RPCs ease the formulation of multi-sensor block adjustment for SAR-optical imagery.

The block adjustment is used to align the optical imagery with respect to the SAR data. Generally, the absolute geolocalization accuracy of optical satellite imagery is lower than that of modern SAR sensors. Evaluations show that the absolute accuracy of geopositioning using TerraSAR-X imagery is within a single resolution cell in both the azimuth and range directions, and can even go down to the cm-level (Eineder et al., 2011). In contrast, the absolute accuracy of geolocalization using basic WorldView-2 products is generally no better than 3 m (DigitalGlobe, 2018). Consequently, the block adjustment propagates the high geometrical accuracy of SAR data into the final 3D product, thus avoiding the need for external control points.

The main stage of SAR-optical stereogrammetry, however, is a dense matching algorithm for 3D reconstruction (Bagheri et al., 2018). In this study, we use the semi-global matching (SGM) method, which incorporates both mutual information and census, as well as their weighted sum as cost functions, in its core.

The remainder of this paper is organized as follows. First, the modeling of SAR sensor geometries with RPCs is explained in Section 2.1. After briefly introducing the epipolarity constraint and its benefits, a mathematical proof of this constraint for SAR-optical image pairs is presented in Section 2.2. In Section 2.3, the application of multi-sensor block adjustment using RPCs for SAR-optical image pairs is introduced. The principle of the SGM algorithm is recapitulated in Section 2.4. Section 3 summarizes experiments and results of our implementation of the SAR-optical stereogrammetry workflow for TerraSAR-X/WorldView-2 image pairs over two urban study areas. Based on these results, the feasibility of stereogrammetric 3D reconstruction from SAR-optical image pairs over urban areas, as well as its advantages and limitations, are discussed in Section 4. Finally, Section 5 presents the conclusions to this study.

2. SAR-optical stereogrammetry

Fig. 1 shows the general framework of SAR-optical stereogrammetric 3D reconstruction. Similar to optical stereogrammetry, one grayscale optical image and one amplitude SAR image form a stereo image pair that can be processed by suitable matching methods to find all possible conjugate pixels. However, some important pre-processing steps are required before the matching and 3D reconstruction. Currently, most VHR optical images are delivered using RPCs. Thus, the primary step in the SAR-optical stereogrammetry framework is to estimate the RPCs for SAR imagery as well. This process homogenizes the geometry models of both sensors and simplifies the subsequent processes of SAR-optical block adjustment and establishing an epipolarity constraint. The next phase is to carry out multi-sensor block adjustment to align the optical image to the SAR image. This rectifies the RPCs of the optical imagery with respect to the SAR imagery, thus improving the absolute geolocalization of the optical imagery and correcting the positions of the epipolar curves on the optical imagery. A disparity map is then produced in the frame of the reference image via a dense image matching algorithm such as SGM. From this map, the 3D positions of the points can be determined by reconstructing the geometry of the SAR and optical imagery for a particular exposure. However, the success of the aforementioned framework relies on the possibility of establishing an epipolarity constraint for SAR-optical image pairs. Thus, the existence of the epipolarity constraint for SAR-optical image pairs must be investigated. In the following, the details of each step of the SAR-optical stereogrammetry framework are explained and the potential of using an epipolarity constraint for SAR-optical image pairs is investigated.

2.1. Preparation: RPCs for SAR imagery

RPCs are a well-established substitute for the rigorously derived optical imaging model. They are widely used for different purposes such as epipolar curve reconstruction (Oh et al., 2010), block adjustment (Grodecki and Dial, 2003), space resection-intersection and 3D reconstruction (Fraser et al., 2006; Li et al., 2007; Tao and Hu, 2002; Tao et al., 2004; Toutin, 2006) or image rectification (Tao and Hu, 2001). The relation between the image space and the geographic reference system is created by the rational functions (Grodecki et al., 2004)

$$c = \frac{P_1(\lambda, \phi, h)}{P_2(\lambda, \phi, h)} = f(\lambda, \phi, h) \tag{1}$$

and

$$r = \frac{P_3(\lambda, \phi, h)}{P_4(\lambda, \phi, h)} = g(\lambda, \phi, h), \tag{2}$$

where r, c are normalized image coordinates, i.e. normalized rows and columns of points in the scene and ϕ, λ, and h denote the normalized latitude, longitude, and height of the respective ground point. The relationship between normalized and un-normalized coordinates is given by Tao and Hu (2001)

$$X = \frac{X_u - X_o}{S_x}, \tag{3}$$

where X is the normalized coordinate, X_u is the un-normalized value of the coordinate, and X_o, S_x are the offset and scale factors, respectively.

In Eqs. (1) and (2), P_i $(i = 1, \ldots, 4)$ are n-order polynomial functions that are used to model the relationship between the image space and the reference system. They can be written as

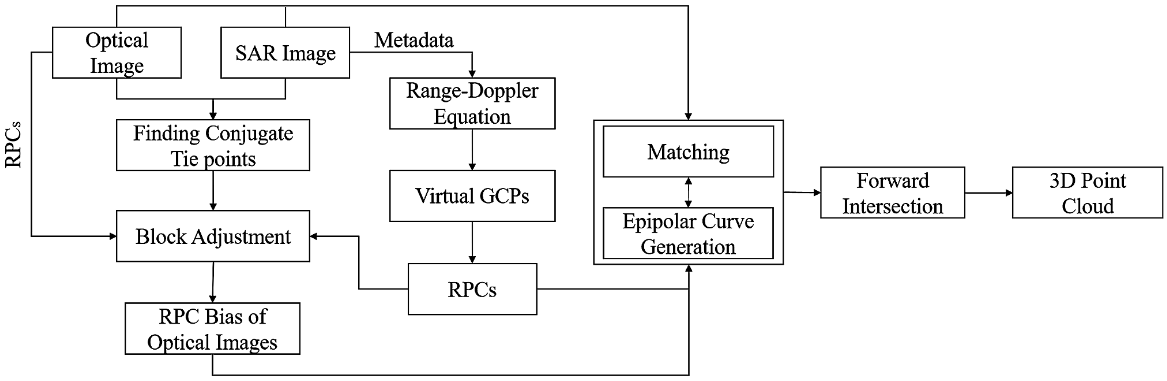


Fig. 1. Framework for stereogrammetric 3D reconstruction from SAR-optical image pairs.

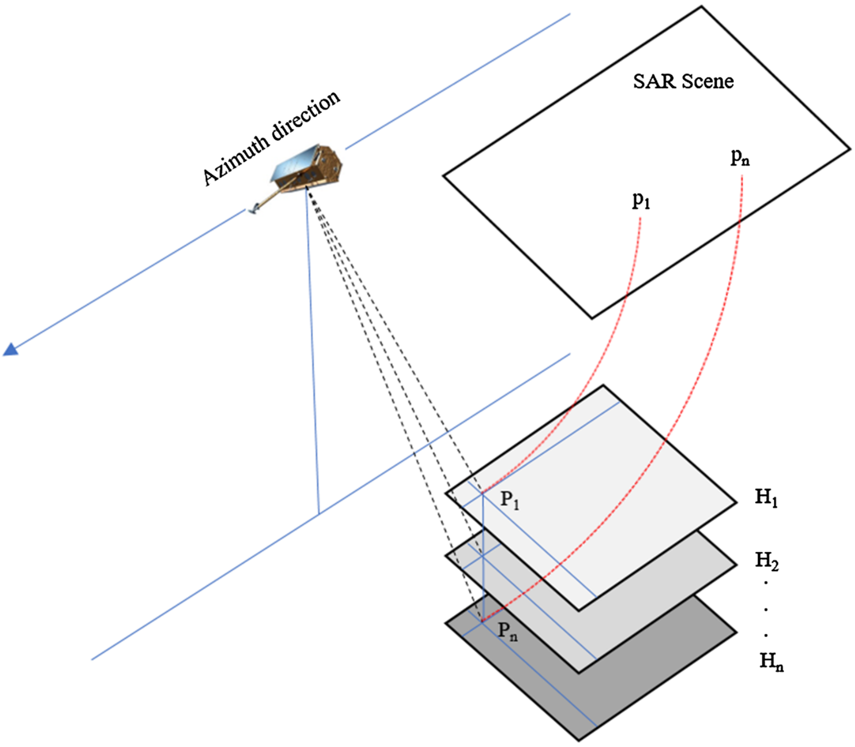


Fig. 2. Procedure of estimating RPCs by terrain-independent approach.

$$P_i = a_{i,0} + a_{i,1}h + a_{i,2}\phi + a_{i,3}\lambda + a_{i,4}h\phi + a_{i,5}h\lambda + a_{i,6}\phi\lambda + a_{i,7}h^2 + a_{i,8}\phi^2 + a_{i,9}\lambda^2 + a_{i,10}h\phi\lambda + a_{i,11}h^2\phi + a_{i,12}h^2\lambda + a_{i,13}\phi^2 h + a_{i,14}\phi^2\lambda + a_{i,15}h\lambda^2 + a_{i,16}\phi\lambda^2 + a_{i,17}h^3 + a_{i,18}\phi^3 + a_{i,19}\lambda^3, \quad (4)$$

where $a_{i,n}$ ($n = 0, 1, \ldots, 19$) are the polynomial coefficients.

For projection from the image space to terrain, the inverse form of the rational function models is used:

$$\lambda = \frac{P_5(c, r, h)}{P_6(c, r, h)} = f'(c, r, h) \quad (5)$$

and

$$\phi = \frac{P_7(c, r, h)}{P_8(c, r, h)} = g'(c, r, h) \quad (6)$$

For this task, another set of RPCs for inverse projection as well as the terrain height h is needed.

The main reason for using RPCs is to facilitate the computational process of the subsequent processing tasks. Instead of describing the stereogrammetric intersection with a combination of the range-Doppler model for the SAR image and a push-broom model for the optical image, from a mathematical point of view the RPC formulation homogenizes everything to a comparably simple joint model. However, fitting RPCs to a sensor model is challenging in its own way and demands sufficient, well-distributed control points. RPCs are usually calculated with either a terrain-independent or terrain-dependent approach (Tao and Hu, 2001). In the terrain-dependent approach, accurate Ground Control Points (GCPs) are used to estimate the RPCs. Thus, the final accuracy of the RPCs depends on the number, accuracy, and distribution of GCPs.

While the terrain-dependent approach is an expensive way of estimating RPCs (and GCPs may not be available for every study area), the terrain-independent method allows RPCs to be estimated without any GCPs (Tao and Hu, 2001). Instead, a set of virtual GCPs (VGCPs), which are related to the image through the rigorous imaging model of the respective sensor, are used to approximate the RPCs. The VGCPs are arranged in a grid-shape format on planes located at different heights over the study area, such as depicted in Fig. 2. The resulting cube of points is then projected to the image space, and their corresponding image coordinates are determined by reconstructing the rigorous model. The RPCs can subsequently be estimated using a least-squares calculation. Note that, when using higher-order RPCs, the least-squares estimation suffers from an ill-posed configuration that causes the results to deviate from their optimal values. In these circumstances, a regularization approach based on Tikhonovs method can be employed to obtain acceptable solutions (Tikhonov and Arsenin, 1977).

The RPCs for optical sensors are usually delivered by vendors alongside the image files. For SAR sensors – with the exception of the Chinese satellite GaoFen-3 – however usually only ephemerids and orbital parameters are attached to the data. Therefore, Zhang et al. investigated the generation of RPCs for various SAR sensors based on the terrain-independent approach (Zhang et al., 2011). Their results show that RPCs can be used as substitutes for the range-Doppler equations with acceptable accuracy. In this study, we use this terrain-independent approach for SAR RPC generation.

2.2. Epipolarity constraint for SAR-optical image matching

In most stereogrammetric 3D reconstruction scenarios, the epipolarity constraint facilitates the procedure of image matching by reducing the search space from 2D to 1D (Morgan et al., 2004). The epipolarity constraint always exists for optical stereo images captured by frame-type cameras that follow a perspective projection (Cho et al., 1993). This phenomenon is illustrated in Fig. 3.

For a point p in the *left-hand* image, the conjugate point in the corresponding *right-hand* image is located on the so-called epipolar line. This epipolar line lies on the plane passing through both image

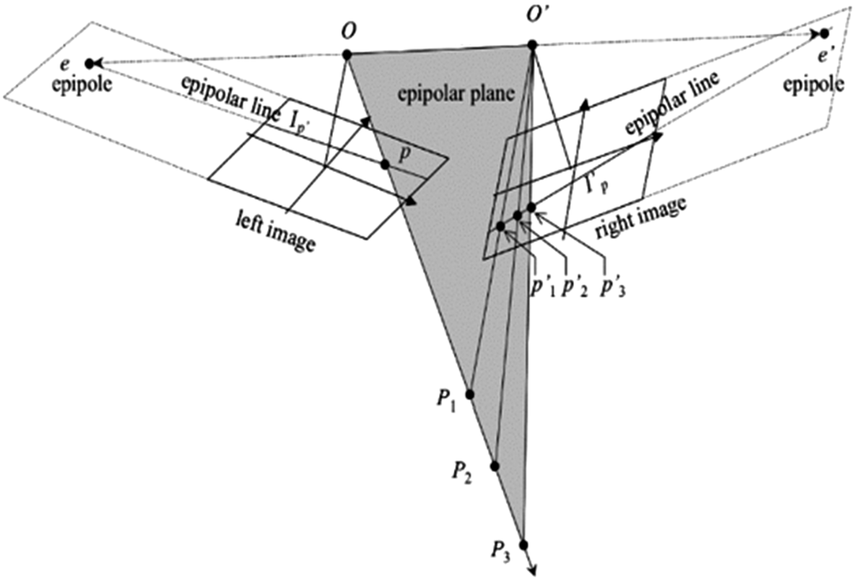


Fig. 3. Epipolarity constraint for frame-type camera (Morgan et al., 2004).

projection centers (O, O') and the image point p. It can also be obtained by changing the depth or height of p in the reference coordinate system. While it is known that epipolar lines exist for images captured from frame-type cameras, straightness cannot be ensured for other sensor types (Cho et al., 1993). We thus refer to epipolar curves instead of epipolar lines to express generality in the remainder of this paper.

With respect to remote sensing, several studies have demonstrated that the epipolar curves for scenes acquired by linear array push-broom sensors are not straight (Gupta and Hartley, 1997; Kim, 2000). For example, Kim (2000) used the model developed by Orun (1994) to prove that the epipolar curves in SPOT scenes looks like hyperbolas. Orun and Natarajan's model assumes that the rotational roll and pitch parameters are constant during the flight, while the yaw can be modeled by quadratic time-dependent polynomials. Morgan et al. (2004) demonstrated that the epipolar curves would not be straight even with uniform motion.

For a SAR sensor, the imaging geometry is completely different from that of optical sensors, as data are collected in a side-looking manner based on the range-Doppler geometry (Curlander, 1982). However, the possibility of establishing the epipolarity constraint in stereo SAR image pairs has been investigated by Gutjahr et al. (2014) and Li and Zhang (2013) for radargrammetric 3D reconstruction. Gutjahr et al. experimentally showed that epipolar curves in SAR image pairs are also not perfectly straight, but can be approximately assumed to be straight for radargrammetric 3D reconstruction tasks through dense matching (Gutjahr et al., 2014).

In this research, we investigate the epipolarity constraint mathematically and experimentally for the unconventional multi-sensor situation of SAR-optical image pairs. In general, epipolar curves in image pairs captured by frame-type cameras (as shown in Fig. 3) can be described as (Hartley and Zisserman, 2004)

$$l_r = \mathbf{F}^T p' \tag{7}$$

where l_r refers to the epipolar curve in the right-hand image associated with the image point p' on the left-hand image. $\mathbf{F}$ is the fundamental matrix, which includes interior and exterior orientation parameters for projecting coordinates between the two images. Similarly, an epipolar curve in the left-hand image can be written as $l_l = \mathbf{F}p''$. For push-broom satellite image pairs, the epipolarity constraint can be verified in a similar way, but linear arrays are substituted for a frame image. Furthermore, the fundamental matrix for push-broom sensors is more complex than that for frame-type sensors. In the following, inspired by the mathematical proof of the epipolarity constraint for stereo optical imagery given in Morgan et al. (2004) and using the epipolar curve equation presented in (7), a rigorous epipolar model for SAR-optical image pairs acquired by space-borne platforms will be constructed. For this task, the optical image is considered as the left-hand image and the SAR image is the right-hand image. Fig. 4 shows the configuration of the SAR-optical stereo case. The points o and s mark the positions of the optical linear array push-broom sensor and the SAR sensor, respectively. Using a collinearity condition, a rigorous model for reconstructing the imaging geometry of linear array push-broom sensors can be expressed as (Kratky, 1989)

$$\begin{pmatrix} x_l = 0 \\ y_l \\ f \end{pmatrix} = \lambda R_{\omega(t)\phi(t)\kappa(t)} \begin{pmatrix} X - X^o(t) \\ Y - Y^o(t) \\ Z - Z^o(t) \end{pmatrix}, \tag{8}$$

where (x_l, y_l) are the coordinates of point p in the linear array coordinate system, f is the focal length, $(X^o(t), Y^o(t), Z^o(t))$ represents the satellite position at time t in the reference coordinate system, (X, Y, Z) are the ground coordinates of the target point T, λ is the scale factor, and $\mathbf{R}_{\omega(t)\phi(t)\kappa(t)}$ is the 3D rotational matrix computed from rotations $\omega(t)$, $\phi(t)$, $\kappa(t)$ along the three dimensions at time t. Note that the aforementioned rotational and translational components are estimated by time-dependent polynomials.

A rigorous model based on the range-Doppler geometry (Curlander and McDonough, 1991) (displayed in Fig. 4 as well) can also be applied to the SAR imagery. In this model, the slant-range equation is first used to describe the range sphere as (Curlander, 1982):

$$R = \|\mathbf{R}_{CT} - \mathbf{R}_{CS}\| \tag{9}$$

where R is the slant-range and $\mathbf{R}_{CT}$, $\mathbf{R}_{CS}$ are the target point and SAR sensor position vectors in the reference coordinate system. C refers to the center of the reference coordinate system.

For a given pixel y_r in the slant-range SAR scene, Eq. (9) can be reformulated as

$$R = c\ t = c\ (t_0 + \frac{y_r}{2f_r}) = c\ t_0 + c\ \frac{y_r}{2f_r} = R_0 + \Gamma\, y_r, \tag{10}$$

where R is the slant-range of the target point, c is the velocity of light,

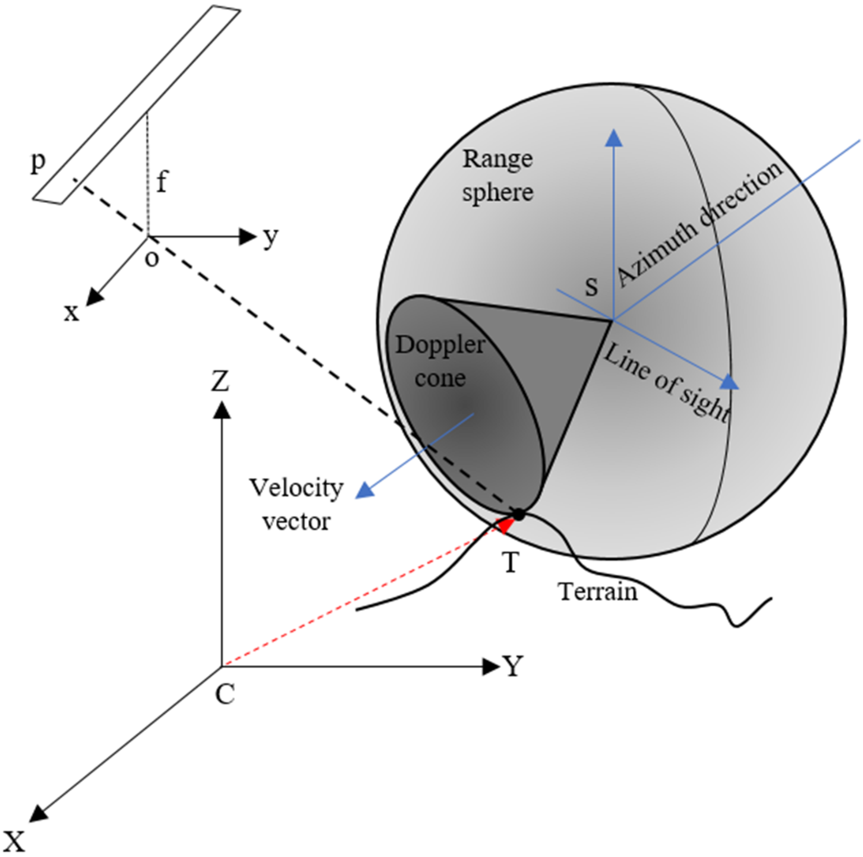


Fig. 4. Imaging geometry for configuration of SAR-optical imagery.

t_0, t are the one-way signal transmission times for the first range pixel and range pixel coordinate y_r, respectively, and f_r is the range sampling rate. R_0 gives the slant-range for the first range pixel and $\Gamma = \frac{c}{2f_r}$.

The second equation describes the geometry of the Doppler cone:

$$f_D = \frac{2}{\lambda_r R}\mathbf{V}\cdot(\mathbf{R}_{CT} - \mathbf{R}_{CS}), \tag{11}$$

where f_D is the Doppler frequency, λ_r is the SAR signal wavelength, $\mathbf{V}$ is the velocity vector and $\cdot$ denotes the inner product operator.

For the epipolarity constraint, we assume that the target point T is imaged at time τ by the push-broom sensor. If we fix the time variable t with τ and consider the corresponding image coordinate in the linear array coordinate system as $(0, y_l, f)$, we can use Eq. (8) to back-project from the linear array coordinate system to the terrestrial reference system as follows:

$$\begin{pmatrix} X - X^o \\ Y - Y^o \\ Z - Z^o \end{pmatrix} = \lambda^{-1} M_{\omega\phi\kappa} \begin{pmatrix} 0 \\ y_l \\ f \end{pmatrix} \tag{12}$$

where $M_{\omega\phi\kappa} = R^T_{\omega\phi\kappa}$. For imaging time t, all time-dependent parameters are estimated under the constraint $t = \tau$. Thus, for this specific instance, the variable index t is eliminated from Eq. (8). By expanding Eq. (12) and removing the scale factor effect, we have:

$$X - X^o = (Z - Z^o)\frac{m_{11}\,0 + m_{12}\,y_l + m_{13}\,f}{m_{31}\,0 + m_{32}\,y_l + m_{33}\,f} = (Z - Z^o)\frac{m_{12}\,y_l + m_{13}\,f}{m_{32}\,y_l + m_{33}\,f} \tag{13}$$

$$Y - Y^o = (Z - Z^o)\frac{m_{21}\,0 + m_{22}\,y_l + m_{23}\,f}{m_{31}\,0 + m_{32}\,y_l + m_{33}\,f} = (Z - Z^o)\frac{m_{22}\,y_l + m_{23}\,f}{m_{32}\,y_l + m_{33}\,f} \tag{14}$$

where m_{ij} are elements of matrix $M_{\omega\phi\kappa}$.

If the velocity vector of the SAR sensor is computed in the zero-Doppler frequency transition, we can reformulate Eq. (11) as:

$$V_x(t)(X - X^s(t)) + V_y(t)(Y - Y^s(t)) + V_z(t)(Z - Z^s(t)) = 0 \tag{15}$$

where $(V_x(t), V_y(t), V_z(t))$ are the components of velocity vector $\mathbf{V}$. Hence:

$$V_x(t)X - V_x(t)X^s(t) + V_y(t)Y - V_y(t)Y^s(t) + V_z(t)Z - V_z(t)Z^s(t) = 0 \tag{16}$$

From Eqs. (13) and (14), we can derive.

$$X - \left(\frac{m_{12}\,y_l + m_{13}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z = X^o - \left(\frac{m_{12}\,y_l + m_{13}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z^o \tag{17}$$

$$Y - \left(\frac{m_{22}\,y_l + m_{23}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z = Y^o - \left(\frac{m_{22}\,y_l + m_{23}\,f}{m_{32}\,y_l + m_{33}\,f}\right)Z^o \tag{18}$$

Multiplying both sides of the Eqs. (17) and (18) by $-V_x(t)$ and $-V_y(t)$, respectively, and then combining them with Eq. (16), Z can be calculated as follows:

$$Z = \frac{(m_{32}\, y_l + m_{33}\, f)\,[V_x(t)(X^s(t) - X^o) + V_y(t)(Y^s(t) - Y^o) + V_z(t)Z^s(t)]}{(m_{12}\, y_l + m_{13}\, f)\, V_x(t) + (m_{22}\, y_l + m_{23}\, f)\, V_y(t) + (m_{32}\, y_l + m_{33}\, f)\, V_z(t)} + \frac{[(m_{12}\, y_l + m_{13}\, f)\, V_x(t) + (m_{22}\, y_l + m_{23}\, f)\, V_y(t)]Z^o}{(m_{12}\, y_l + m_{13}\, f)\, V_x(t) + (m_{22}\, y_l + m_{23}\, f)\, V_y(t) + (m_{32}\, y_l + m_{33}\, f)\, V_z(t)} \tag{19}$$

Changing the position of target point T in the Z direction is equivalent to changing the corresponding image coordinates on the epipolar curve. Consequently, the determination of image coordinates in the SAR scene can be realized by tracking the sensor positions in the respective instances. In fact, in spite of fixing the position of the target point for the optical sensor at time τ, the location components $(X^s(t), Y^s(t), Z^s(t))$ and the velocity components of the SAR sensor $(V_x(t), V_y(t), V_z(t))$ are time-dependent and can be estimated for each instant using time-dependent polynomials. For the sake of simplicity, the SAR sensor trajectory can be approximated by a linear motion model. Thus, the velocity components $(V_x(t), V_y(t), V_z(t))$ will remain constant with time at $(X^s(t), Y^s(t), Z^s(t))$ i.e., the acceleration is 0 and the location components $(X^s(t), Y^s(t), Z^s(t))$ can be calculated using linear time-dependent functions:

$$\begin{aligned} X^s(t_a) &= X_0^s + V_x\, t_a \\ Y^s(t_a) &= Y_0^s + V_y\, t_a \\ Z^s(t_a) &= Z_0^s + V_z\, t_a \end{aligned} \tag{20}$$

where (X_0^s, Y_0^s, Z_0^s) is the position of the SAR sensor at initial time t_0. The time t_a is the azimuthal time, which can be expressed according to the line coordinate x_r in the SAR scene as

$$t_a = \frac{x_r}{PRF} = k\, x_r, \tag{21}$$

where PRF is the pulse repetition frequency in Hz.

Substituting the parameters expressed in Eqs. (20) and (21) into (19), we obtain:

$$\begin{aligned} Z = &\frac{(m_{32}\, y_l + m_{33}\, f)\,[V_x(X_0^s - X^o) + V_y(Y_0^s - Y^o) + V_z Z_0^s]}{(m_{12}\, y_l + m_{13}\, f)\, V_x + (m_{22}\, y_l + m_{23}\, f)\, V_y + (m_{32}\, y_l + m_{33}\, f)\, V_z} \\ &+ \frac{[(m_{12}\, y_l + m_{13}\, f)\, V_x + (m_{22}\, y_l + m_{23}\, f)\, V_y]Z^o}{(m_{12}\, y_l + m_{13}\, f)\, V_x + (m_{22}\, y_l + m_{23}\, f)\, V_y + (m_{32}\, y_l + m_{33}\, f)\, V_z} \\ &+ k\, x_r \\ &\left(\frac{V_x^2 + V_y^2 + V_z^2}{(m_{12}\, y_l + m_{13}\, f)\, V_x + (m_{22}\, y_l + m_{23}\, f)\, V_y + (m_{32}\, y_l + m_{33}\, f)\, V_z}\right) \end{aligned} \tag{22}$$

Eq. (22) can be simplified to:

$$Z = c_0 + c_1\, x_r \tag{23}$$

and substituting (23) into (17) and (18) gives:

$$\begin{aligned} X = X^o &- \left(\frac{m_{12}\, y_l + m_{13}\, f}{m_{32}\, y_l + m_{33}\, f}\right) Z^o + \left(\frac{m_{12}\, y_l + m_{13}\, f}{m_{32}\, y_l + m_{33}\, f}\right)(c_0 + c_1\, x_r) \\ = X^o &- \left(\frac{m_{12}\, y_l + m_{13}\, f}{m_{32}\, y_l + m_{33}\, f}\right) Z^o + \left(\frac{m_{12}\, y_l + m_{13}\, f}{m_{32}\, y_l + m_{33}\, f}\right) c_0 \\ &+ \left(\frac{m_{12}\, y_l + m_{13}\, f}{m_{32}\, y_l + m_{33}\, f}\right) c_1\, x_r \end{aligned} \tag{24}$$

$$\begin{aligned} Y = Y^o &- \left(\frac{m_{22}\, y_l + m_{23}\, f}{m_{32}\, y_l + m_{33}\, f}\right) Z^o + \left(\frac{m_{22}\, y_l + m_{23}\, f}{m_{32}\, y_l + m_{33}\, f}\right)(c_0 + c_1\, x_r) \\ = Y^o &- \left(\frac{m_{22}\, y_l + m_{23}\, f}{m_{32}\, y_l + m_{33}\, f}\right) Z^o + \left(\frac{m_{22}\, y_l + m_{23}\, f}{m_{32}\, y_l + m_{33}\, f}\right) c_0 \\ &+ \left(\frac{m_{22}\, y_l + m_{23}\, f}{m_{32}\, y_l + m_{33}\, f}\right) c_1\, x_r \end{aligned} \tag{25}$$

Eqs. (24) and (25) can be written in the form:

$$X = a_0 + a_1\, x_r \tag{26}$$

$$Y = b_0 + b_1\, x_r \tag{27}$$

The slant-range represented by Eq. (10) can be reformulated as:

$$(X - X^s)^2 + (Y - Y^s)^2 + (Z - Z^s)^2 = (R_0 + \Gamma\, y_r)^2 \tag{28}$$

Substituting (26), (27), (23) and (20) into (28), we have:

$$\begin{aligned} &(a_0 + a_1\, x_r - X_0^s - V_x\, k\, x_r)^2 + (b_0 + b_1\, x_r - Y_0^s - V_y\, k\, x_r)^2 \\ &\quad + (c_0 + c_1\, x_r - Z_0^s - V_z\, k\, x_r)^2 = (R_0 + \Gamma\, y_r)^2 \end{aligned} \tag{29}$$

For simplicity, setting $A_0 = a_0 - X_0^s$, $A_1 = a_1 - V_x\, k$, $B_0 = b_0 - Y_0^s$, $B_1 = b_1 - V_y\, k$, $C_0 = c_0 - Z_0^s$, $C_1 = c_1 - V_z\, k$ gives:

$$(A_0 + A_1\, x_r)^2 + (B_0 + B_1\, x_r)^2 + (C_0 + C_1\, x_r)^2 = (R_0 + \Gamma\, y_r)^2 \tag{30}$$

which can be expanded to yield:

$$\begin{aligned} &(A_0^2 + B_0^2 + C_0^2) + 2(A_0\, A_1 + B_0\, B_1 + C_0\, C_1)\, x_r + (A_1^2 + B_1^2 + C_1^2)\, x_r^2 \\ &\quad = (R_0 + \Gamma\, y_r)^2 \end{aligned} \tag{31}$$

If $A_0^2 + B_0^2 + C_0^2 = F_0$, $2(A_0 A_1 + B_0 B_1 + C_0 C_1) = F_1$, and $A_1^2 + B_1^2 + C_1^2 = F_2$, this can be rewritten as:

$$\Gamma\, y_r = \sqrt{F_2\, x_r^2 + F_1\, x_r + F_0} - R_0 \tag{32}$$

Eq. (32) is a general rigorous model representing the epipolarity constraint for SAR-optical image pairs based on their imaging parameters contained in F_0, F_1, F_2, Γ and R_0. This shows that an epipolarity-like constraint can be established for SAR-optical image pairs. However, the non-linear relation between y_r and x_r in Eq. (32) shows that SAR-optical epipolar curves are not straight, even under the assumption of linear motion for the SAR system. In Section 3.3, this epipolarity constraint will be experimentally investigated for an RPC-based imaging model.

2.3. SAR-optical multi-sensor block adjustment

As illustrated in Fig. 1, the main step before implementing dense image matching is to align the optical image to the SAR image. This process is performed using a multi-sensor block adjustment which is based on RPCs instead of rigorous sensor models as proposed in Grodecki and Dial (2003). The block adjustment process improves the relative orientation between both images fixed to the more accurate SAR image orientation parameters. Through the block adjustment, the bias components induced by attitude, ephemeris, and drift errors in the optical image are compensated (d'Angelo and Reinartz, 2012).

The main bias compensation for the RPCs of the optical image involves translating the locations of the epipolar curves to accurate positions using the SAR geopositioning accuracy. Generally, designing an appropriate function for modeling the existing bias in the RPCs given by the optical image depends on the sensor properties (Tong et al., 2010), but for most sensors an affine model can be applied (Fraser and Hanley, 2005). Even for the current generation of VHR linear push-broom array sensors such as WorldView-2, employing only the shift parameters will be sufficient. The affine model for RPC bias compensation can be formulated as

$$\begin{aligned} \Delta x &= m_0 + m_1\, x_o + m_2\, y_o \\ \Delta y &= n_0 + n_1\, x_o + n_2\, y_o, \end{aligned} \tag{33}$$

where x_o, y_o represent column and row of tie points in the optical images and m_i and n_i $(i = 0, 1, 2)$ are unknown affine parameters to be estimated through the block adjustment procedure. Note that tie points are the common points between the SAR and optical images, and can be obtained by manual or automatic sparse matching between two images. Since the automation of the tie point generation process is not the focus of this study, we refer the reader to possible solutions described in Suri and Reinartz (2010), Perko et al. (2011), Merkle et al. (2017).

The geographic coordinates of the tie points in the SAR image are calculated by the inverse rational functions computed for the SAR

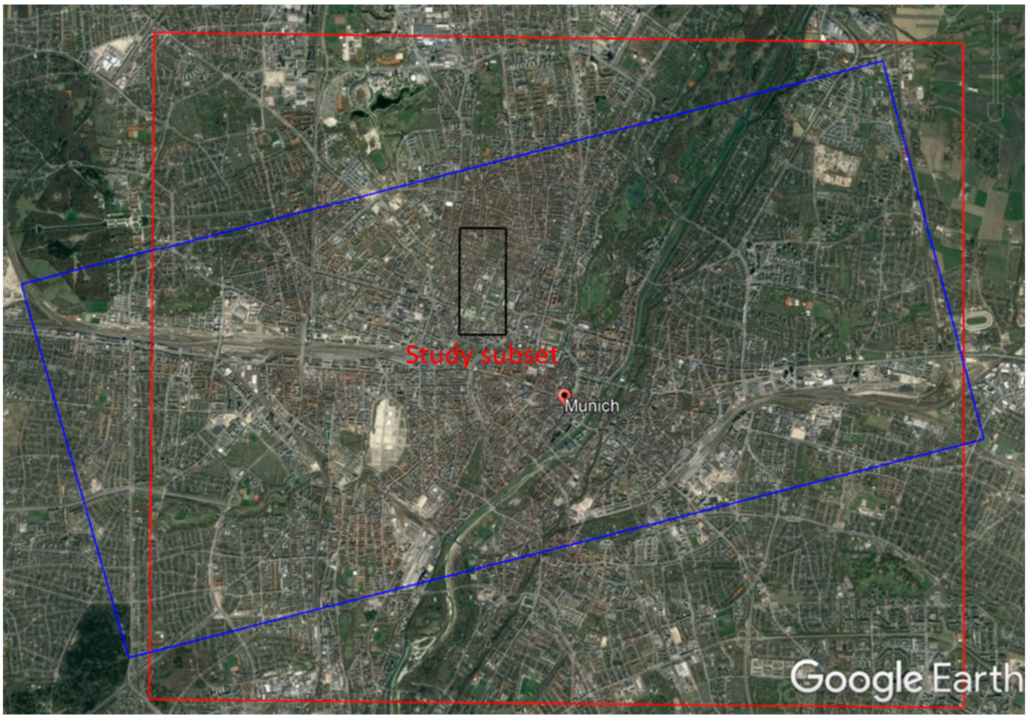


Fig. 5. Display the location of SAR-optical image pairs of the Munich study area. The red and blue rectangles identify areas covered by the WorldView-2 and TerraSAR-X images, respectively, and the black rectangle displays the study subset selected for stereogrammetrix 3D reconstruction over Munich. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

imagery as described in Section 2.1:

$$\lambda^i = f_s'(x_s^i, y_s^i, H) \tag{34}$$

and

$$\phi^i = g_s'(x_s^i, y_s^i, H) \tag{35}$$

where λ^i and ϕ^i are the normalized longitude and latitude of tie point i with normalized image coordinates x_s^i and y_s^i (index s denotes the SAR scene), and here, H is a constant, e.g., the mean height of the study area. f_s' and g_s' are inverse rational functions computed for the SAR sensor to project the tie points from the SAR image to the reference system. The output is a collection of GCPs that can be applied for the RPC rectification of the optical imagery. The resulting GCPs are then projected by the rational function associated with the optical images to give the image coordinates of the GCPs:

$$c_o^i = f_o(\lambda^i, \phi^i, H) \tag{36}$$

and

$$r_o^i = g_o(\lambda^i, \phi^i, H) \tag{37}$$

where c_o^i, r_o^i are the normalized image coordinates of tie point i computed by the forward rational functions of the optical sensor, f_o and g_o.

From Eqs. (33), (36), and (37), the primary equations for SAR-optical block adjustment are formed as:

$$x_o^i = c_{ou}^i + \Delta x^i + v_x^i \tag{38}$$

and

$$y_o^i = r_{ou}^i + \Delta y^i + v_y^i \tag{39}$$

where, x_o^i and y_o^i denote the column and row of tie point i in the optical scene, and c_{ou}^i and r_{ou}^i are the un-normalized coordinates of the tie point after projection and back-projection using the RPCs. The block adjustment equations can then be written as:

$$F_x^i = -x_o^i + c_{ou}^i + \Delta x^i + v_x^i = 0, \tag{40}$$

and

$$F_y^i = -y_o^i + r_{ou}^i + \Delta y^i + v_y^i = 0. \tag{41}$$

Finally, through an iterative least-squares adjustment (Grodecki and Dial, 2003), the unknown parameters m_i and n_i are estimated and the affine model can be formed. This affine model is added to the rational functions of the optical image to improve the geolocation accuracy to that of the SAR image.

2.4. SGM for dense multi-sensor image matching

The core step in a stereogrammetric 3D reconstruction workflow is the dense image matching algorithm to obtain the disparity map, which can then be transformed into the desired 3D point cloud. Generally, there are two different dense matching rationales that can be used according to whether local or global optimization is more important (Brown et al., 2003). For the case of global optimization, an energy functional consisting of two terms is established to find the optimal disparity map (Scharstein et al., 2001):

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d) \tag{42}$$

where $E_{data}(d)$ is a fidelity term that makes the computed disparity map consistent with the input image pairs, $E_{smooth}(d)$ considers the smoothness condition for the disparity map, and λ is a regularization parameter that balances the fidelity and smoothness terms.

For a given image pair, the disparity map is calculated by minimizing the energy functional in (42). The main advantage of global dense matching over local matching methods is greater robustness against noise (Brown et al., 2003), although most existing algorithms for global dense image matching have a greater computational cost

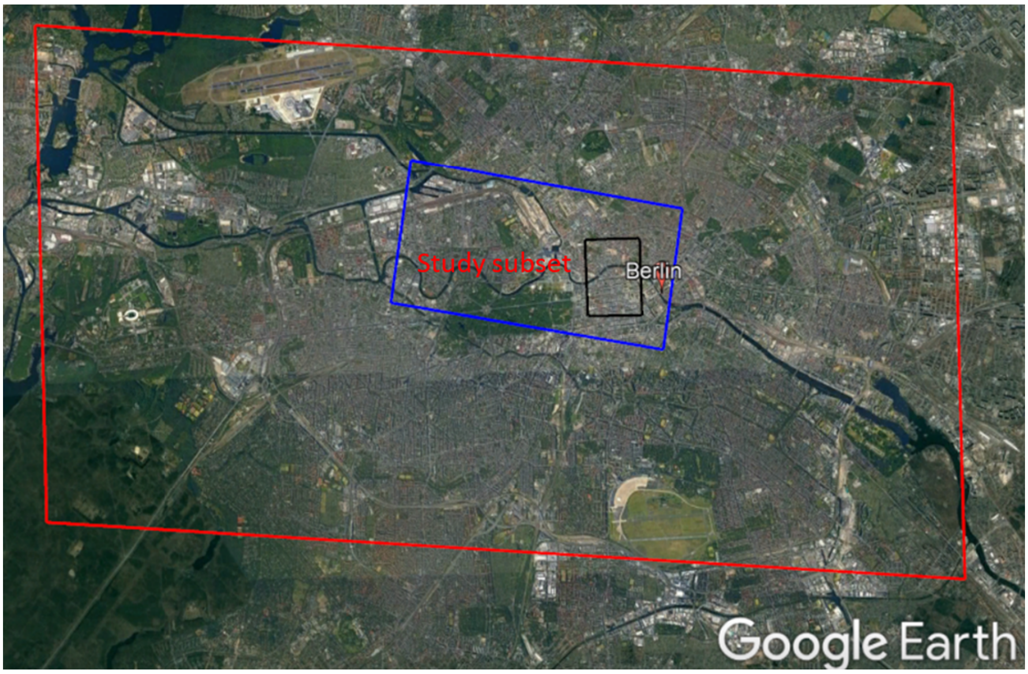


Fig. 6. Display the location of SAR-optical image pairs of the Berlin study area. The red and blue rectangles identify areas covered by the WorldView-2 and TerraSAR-X images, respectively, and the black rectangle displays the study subset selected for stereogrammetrix 3D reconstruction over Berlin. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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

Area	Sensor	Acquisition Mode	Off-Nadir Angle (°)	Ground Pixel Spacing (m)	Acquisition date
Munich	TerraSAR-X WorldView-2	Spotlight Panchromatic	22.99 5.2	0.85 × 0.45 0.5 × 0.5	03.2015 07.2010
Berlin	TerraSAR-X WorldView-2	Staring Spotlight Panchromatic	36.11 29.1	0.17 × 0.45 0.5 × 0.5	04.2016 05.2013

(Hirschmüller, 2008).

For the experiments presented in this paper, we use the well-known SGM method (Hirschmüller, 2008), which offers acceptable computational cost and high efficiency, and performs very similarly to global dense image matching.

2.4.1. Cost functions used in SGM

In this study, the ability of performing dense image matching for SAR-optical image pairs using SGM is investigated. For this purpose, two different cost functions, namely Mutual Information (MI) and Census, as well as their weighted sum, are examined for the dense matching of high-resolution SAR and optical imagery. Typically, the similarity measures employed in the cost function are either signal-based or feature-based metrics (Hassaballah et al., 2016). Classically, signal-based similarity measures such as Normalized Cross Correlation (NCC) and MI are preferable to feature-based similarity measures when used in dense image matching algorithms because of faster calculation.

Among the signal-based matching measures, MI was recommended for SGM as it is known to perform well for images with complicated illumination relationship, such as SAR-optical image pairs (Suri and Reinartz, 2010).

Another similarity measure used in the SGM cost function is Census, which actually acts as a nonparametric transformation. The weighted sum of MI and Census is beneficial for 3D reconstruction in urban areas, especially for reconstructing the footprints of buildings to produce sharper and clearer images (Zhu et al., 2011). The weighted similarity measure can be defined as

$$SM = \alpha\, MI + (1-\alpha)\, \text{Census}, \tag{43}$$

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

2.4.2. SGM settings for efficient SAR-optical dense matching

To increase the efficiency of the SGM performance, some important settings for the dense matching of SAR-optical image pairs must be considered. The basic principle of 3D reconstruction by dense matching is to use the epipolarity constraint to limit the search space. Usually, before dense matching, normal images are created by resampling the original images according to epipolar geometry (Morgan et al., 2004; Oh et al., 2010). In this study, we use the RPC model to realize the SAR-optical epipolar geometry implicitly without the need to generate normal images. This is done by implementing projections and back-projections from the reference image to the ground and back to the corresponding image, respectively, for a specified height range using rational functions. Then, the search for computing disparities can be performed along the thus-created epipolar curves.

In addition, the minimum and maximum disparity values should be selected to restrict the length of the search space along the epipolar

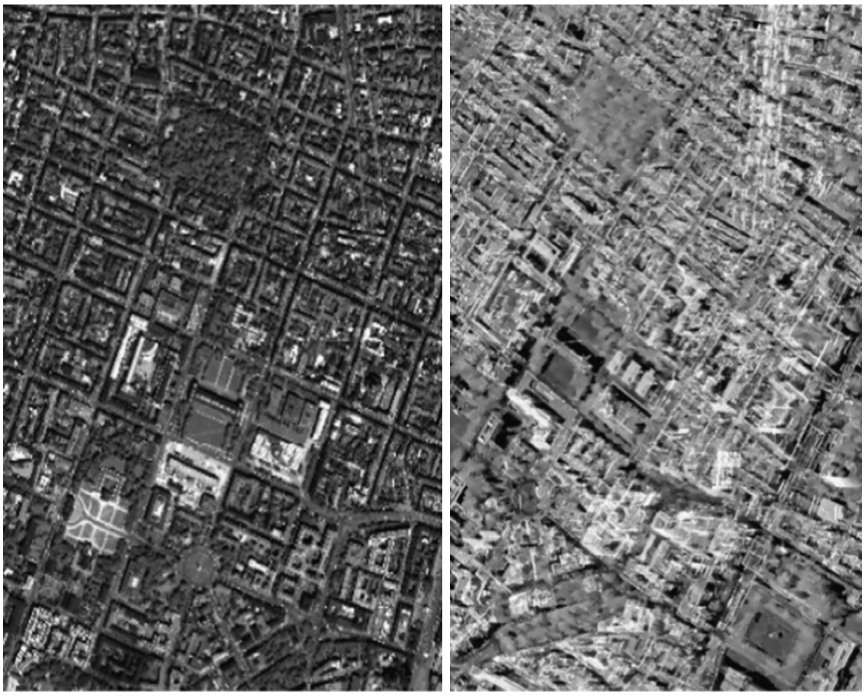


Fig. 7. Display of SAR-optical sub-scenes extracted from Munich study areas (the left-hand image is from WorldView-2, right-hand image is from TerraSAR-X).

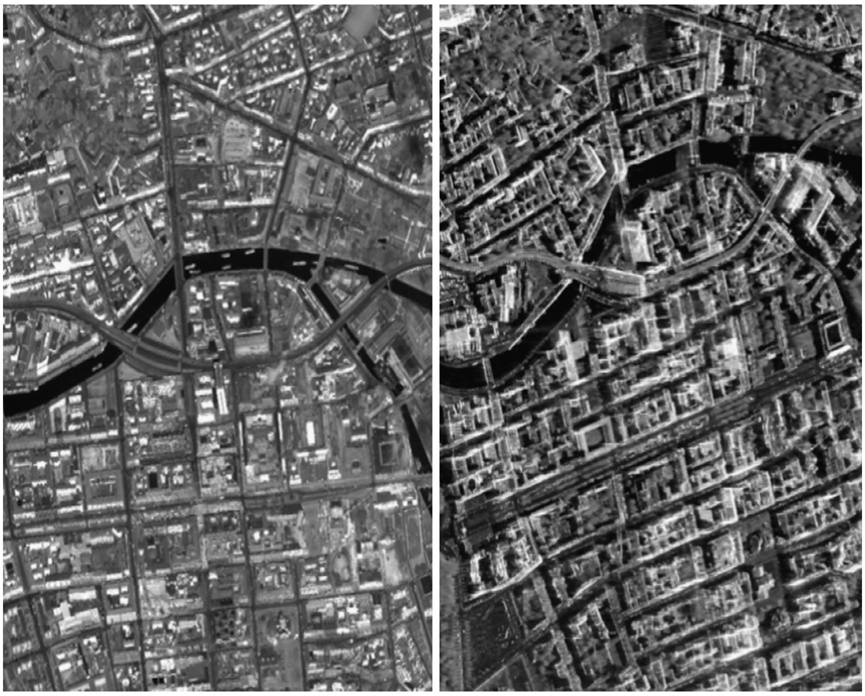


Fig. 8. Display of SAR-optical sub-scenes extracted from Berlin study areas (the left-hand image is from WorldView-2, right-hand image is from TerraSAR-X).

curves. In general, there is more flexibility regarding the selection of this disparity interval for optical image pairs than for SAR-optical image pairs, and using unsuitable values will result in more outliers. The minimum and maximum disparity values can be determined using external data such as the SRTM digital elevation model, which is available for most land surfaces around the world (USGS, 2000). For sake of exploiting the simplicity offered by comparably flat study scenes, we just add and subtract 20-m height differences to the mean terrain heights of the study scenes to obtain the disparity thresholds.

The next setting is to switch off the *minimum region size* option in the

Table 2
Accuracy (standard deviation: STD) of RPCs fitted on SAR sensor model (units: m).

Area	Virtual GCPs		Check points	
	Row	Column	Row	Column
Munich	0.00026	0.00114	0.00025	0.00031
Berlin	0.00024	0.00027	0.00026	0.00118

SGM algorithm, which is usually used to decrease the noise level in the stereogrammetric 3D reconstruction of optical image pairs by eliminating isolated patches from the disparity map based on their small size. Experimental results show that, for SAR-optical image pairs, the complex illumination relationship between the images and the different imaging effects (especially for urban areas) make the *minimum region size* criterion useless, as connectivity cannot be ensured in the disparity map.

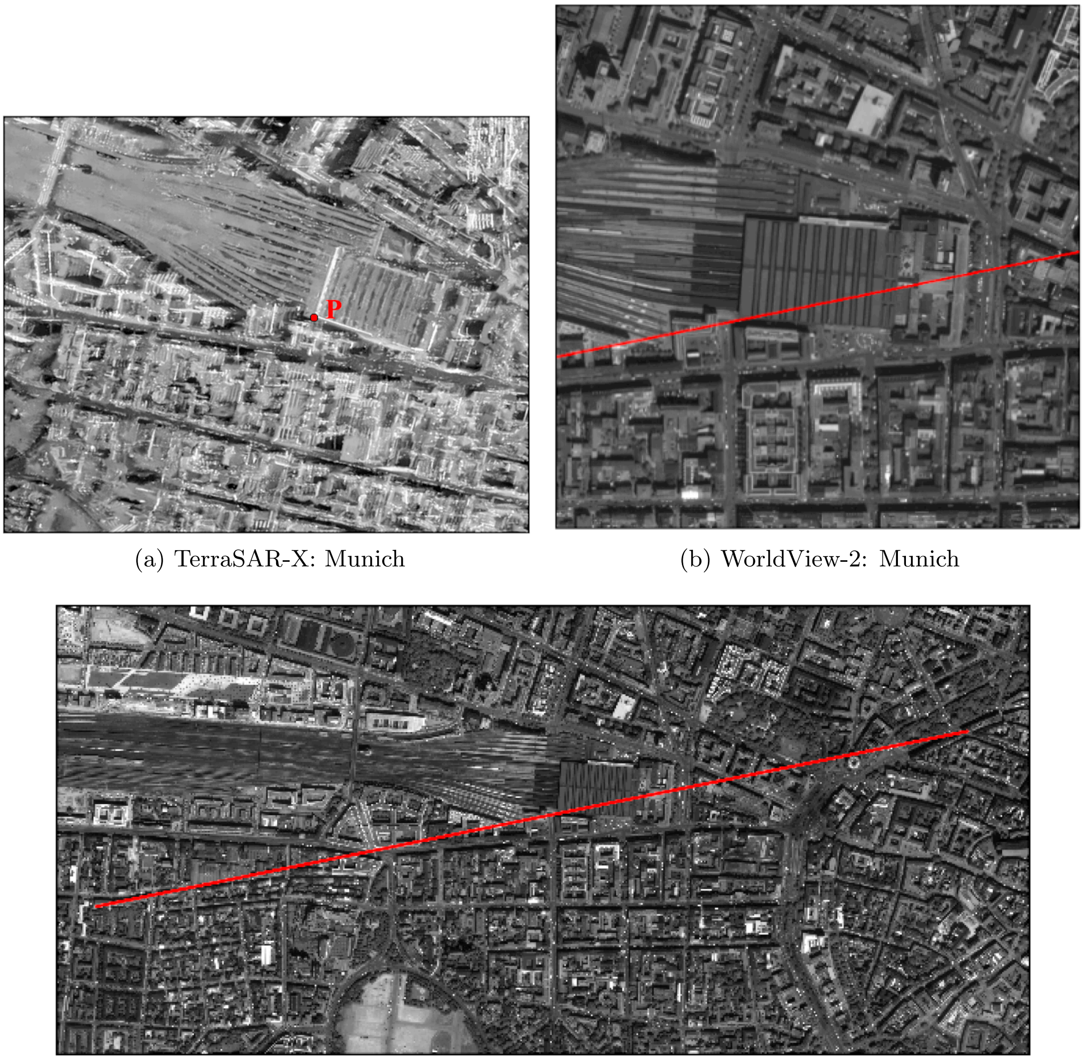
Similar to other dense matching cases, we use the LR (Left-Right) check to investigate binocular half-occlusions (Egnal and Wildes, 2002). This strategy changes the reference images from left to right, and consequently produces two disparity maps that can be checked against each other. To reach sub-pixel accuracy, the disparity in each point is estimated by a quadratic interpolation of neighboring disparities.

In this study, SGM is implemented at four hierarchy levels and the aggregated cost is calculated along 16 directions around each point.

3. Experiments and results

3.1. Study areas and datasets

We selected two study areas, one in Berlin and one in Munich (both located in Germany), to investigate the potential for 3D reconstruction from high-resolution SAR-optical image pairs over urban areas. The locations of TerraSAR-X and WorldView-2 images are displayed in Figs. 5 and 6. The properties of the image pairs for each study area are

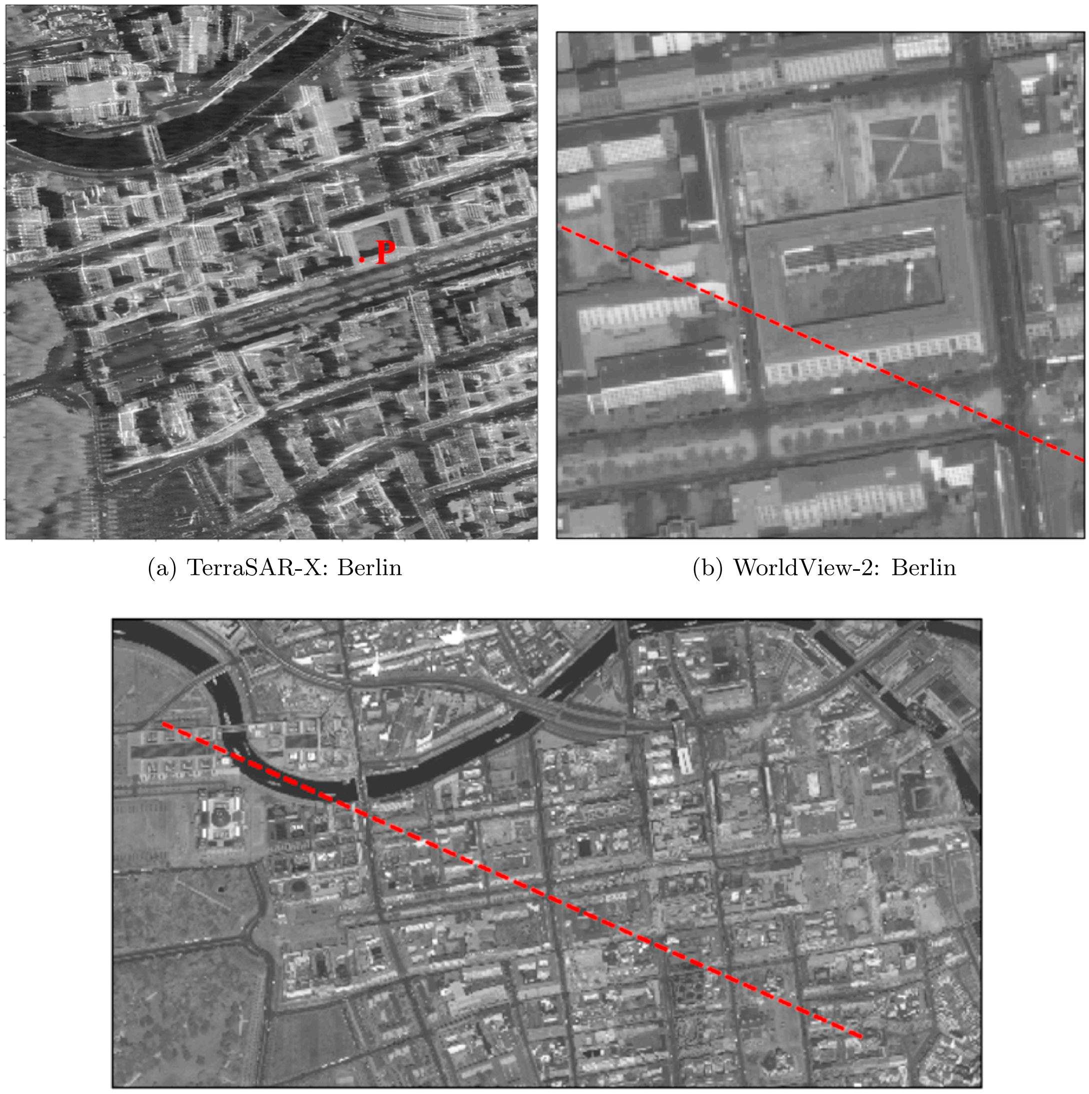


(a) TerraSAR-X: Munich

(b) WorldView-2: Munich

(c) WorldView-2: Munich

Fig. 9. Epipolar curves for the WorldView-2 image (of Munich) given by changing the heights of point p (located at the corner of the Munich central train station in the TerraSAR-X scene) for all possible height values in the image scene. The epipolar curves look like straight, but are not actually straight.



(a) TerraSAR-X: Berlin

(b) WorldView-2: Berlin

(c) WorldView-2: Berlin

Fig. 10. Epipolar curves for the WorldView-2 image (of Berlin) given by changing the heights of point p (a distinct corner of a building in the Berlin TerraSAR-X scene) for all possible height values in the image scene. The epipolar curves look like straight, but are not actually straight.

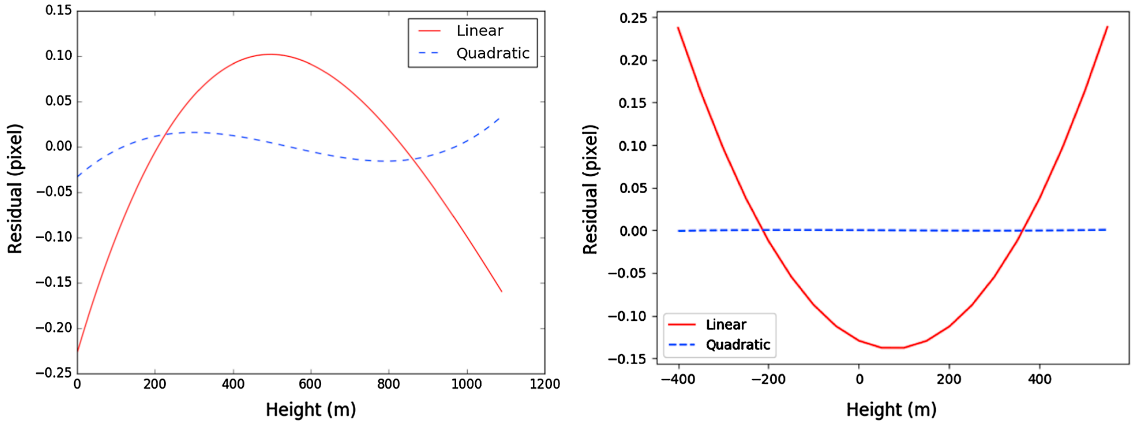
presented in Table 1. In order to enhance the general image similarity and facilitate the matching process, all images were resampled to 1 m × 1 m pixel spacing and the SAR images were filtered with a non-local speckle filter. After implementing bundle adjustment for both datasets, two sub-scenes (with a size of 1000 × 1500 pixels each) from overlapped parts of the study areas were cropped. These sub-scenes are displayed in Figs. 7 and 8.

3.2. Validation of RPCs to model SAR sensor geometry

As described in Section 2.1, RPCs can be used as a substitute for the rigorous range-Doppler model, similar to the standard RPCs delivered with optical imagery. This step is performed to simplify the multi-sensor block adjustment and epipolarity constraint construction. The accuracy of the RPCs can be estimated using independent virtual checkpoints that are produced in a similar way to VGCPs using the range-Doppler model. The word independent implies that the virtual checkpoints are never used in the RPC fitting, i.e., they are located in different positions respective to the VGCPs. The accuracy of the fitted RPCs for the TerraSAR-X data in each study area is listed in Table 2. Analysis was performed based on the residuals of the rows and columns, given by the differences $row_{RPC} - row_{DR}$ and $col_{RPC} - col_{DR}$, i.e., the differences between image coordinates computed by RPCs and range-Doppler. The analysis results confirm that the RPCs can model the range-Doppler geometry for TerraSAR-X data to within a millimeter, and can thus well be used in the 3D reconstruction process.

3.3. Validity of the epipolarity constraint

A general model that proves the epipolarity constraint for SAR-optical image pairs was described in Section 2.2. It was also concluded that epipolar curves are usually not straight. Experimentally, the epipolarity constraint for SAR-optical image pairs can also be modeled based on RPCs. In this paper, we evaluate the epipolarity constraint for



(a) Munich

(b) Berlin

Fig. 11. Linear and quadratic polynomials fitted on the epipolar curves in the WorldView-2 images.

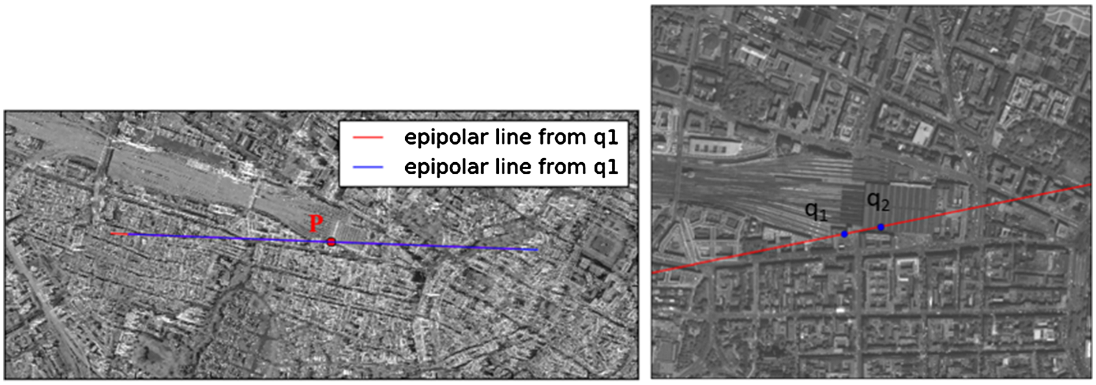
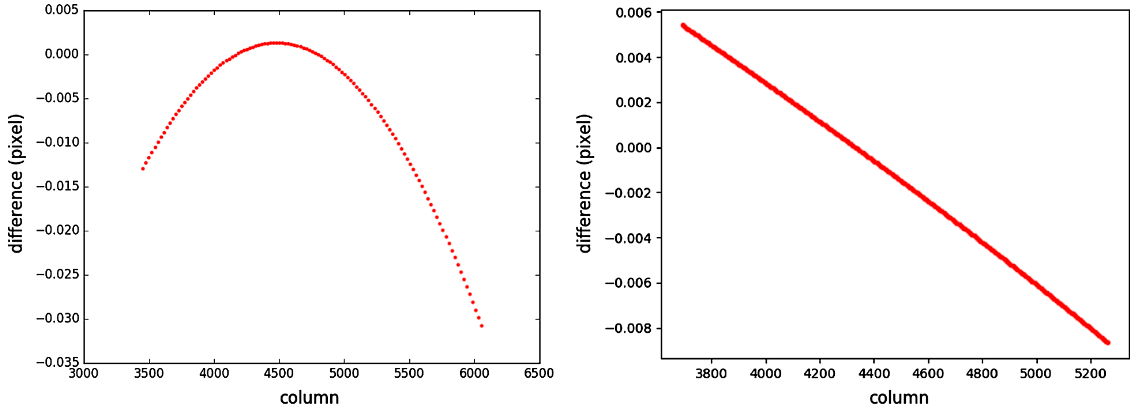


Fig. 12. Corresponding epipolar curves in the Munich TerraSAR-X image (left) derived from two points, q_1 and q_2 on the epipolar curve of the Munich WorldView-2 image (right).



(a) Munich

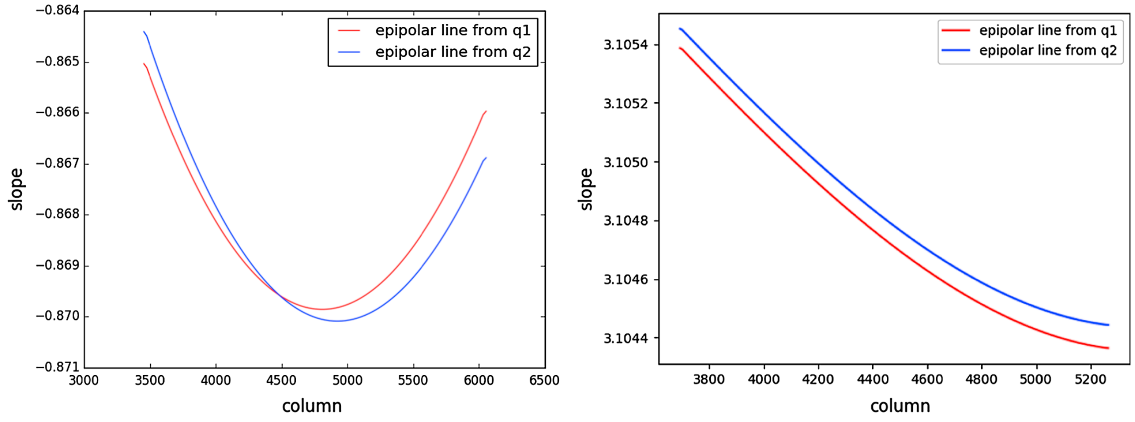
(b) Berlin

Fig. 13. Difference of two corresponding epipolar curves over the column direction. The maximum difference between the two epipolar curves is less than one pixel.

TerraSAR-X and WorldView-2 image pairs acquired over the two study areas.

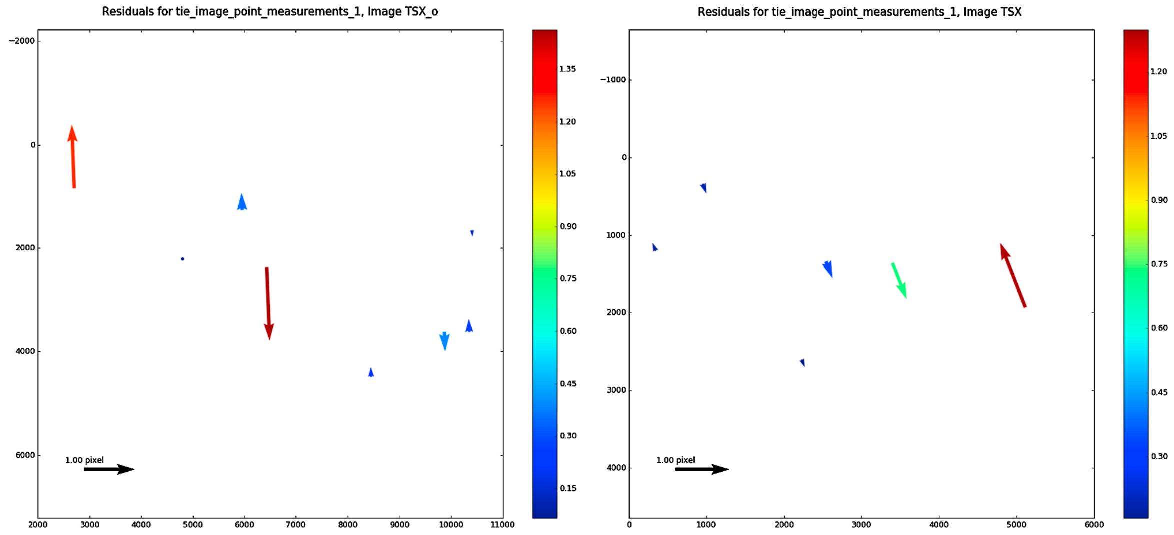
3.3.1. Existence of epipolar curves

We analyzed the validity of the derived SAR-optical epipolarity constraint for an exemplary point located at the corner of the Munich central train station building (p). This point was projected to the terrain space by changing the heights in specific steps, e.g., 10 m, starting from the lowest possible height and proceeding to the highest possible height in the scene (for this experiment we used the interval [0 m, 1200 m]). The output will be an ensemble of points with different heights, such as depicted in Fig. 9(c). All these points were then back-projected to the



(a) Munich (b) Berlin

Fig. 14. Gradients of two epipolar curves constructed from q1 and q2 in TerraSAR-X.



(a) Munich (b) Berlin

Fig. 15. Residuals of tie points after full multi-sensor block adjustment in the TerraSAR-X image space.

Table 3
Block adjustment results (units: m).

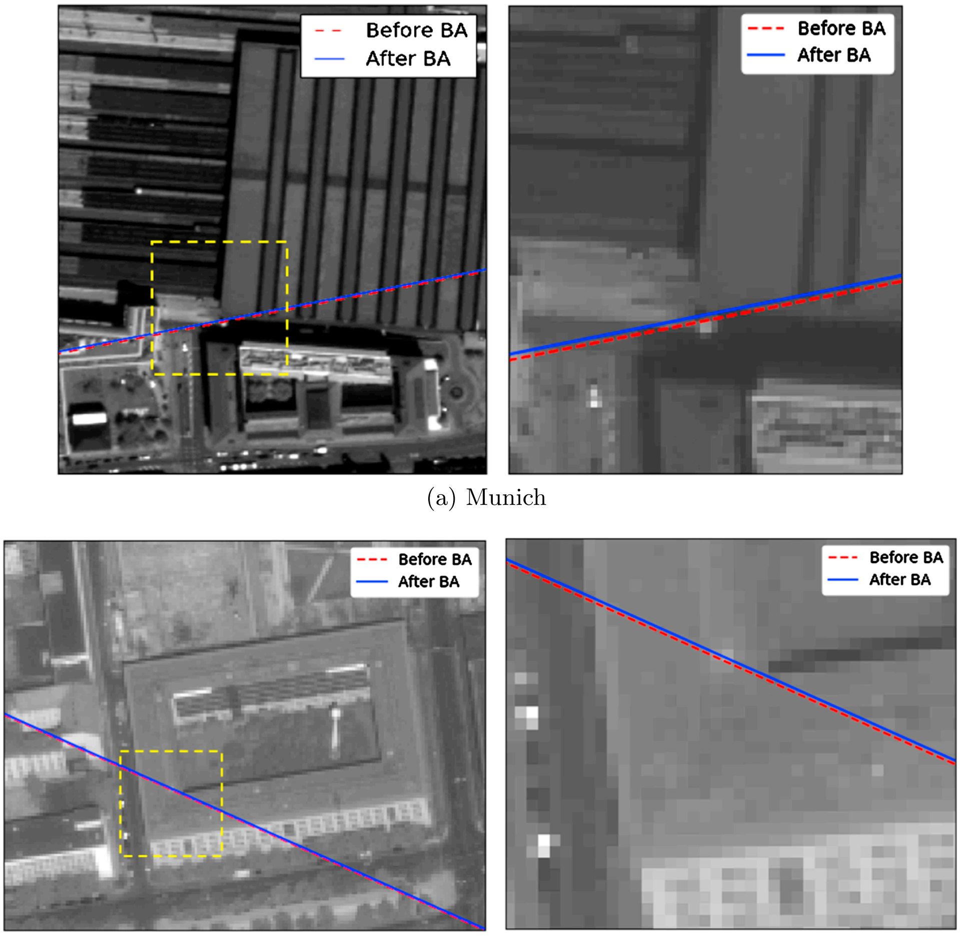
Area	Sensor	Bias Coefficients		STD	MAD	Min	Max	No. of Tie Points
Munich	WorldView-2	− 2.47	− 0.53	0.50	0.14	0.07	1.46	8
	TerraSAR-X	0	0	0.50	0.14	0.07	1.46	
Berlin	WorldView-2	− 0.73	0.28	0.51	0.13	0.19	1.59	6
	TerraSAR-X	0	0	0.42	0.11	0.16	1.30	

WorldView-2 image space using RPCs. The corresponding epipolar curve for all possible heights in the study area is constructed by connecting the image points obtained in this way, as shown in Fig. 9(c). Although the epipolar curve appears to be straight, more analysis is required to determine whether this is the case. By expanding the image, it can be seen in Fig. 9(b) that the epipolar curve nearly passes through the conjugate point of p in the WorldView-2 image. Similarly, another experiment was carried out using the Berlin dataset. The epipolar curve was constructed for an exemplary point located on the corner of a building and for all possible heights in the scene. Fig. 10(d) displays the position of the selected point on the SAR image as well as the corresponding epipolar curve in the WorldView-2 imagery (Figs. 10(e) and (f)).

3.3.2. *Straightness of epipolar curves*

To clarify the straightness of the epipolar curve constructed for point p, linear and quadratic polynomials were fitted to the image points of the epipolar curve. Figs. 11(a) and (b) represent the least-squares residuals with respect to the point heights for the epipolar curves created in both study subsets. The residuals of the linear fit for the epipolar curve established for the Munich WorldView-2 image range from − 0.25 to 0.1 pixels (i.e. meters), whereas the residuals of the quadratic fit are close to zero.

Similar results were given by fitting linear and quadratic polynomials to the image points of the epipolar curve established in the WorldView-2 image of Berlin. Fig. 11(b) clearly shows that the residuals of the epipolar points fitted to the quadratic model are zero, whereas those of the linear fit vary between − 0.15 m and 0.25 m. Both analyses illustrate that the constructed epipolar curves are not straight.



(a) Munich

(b) Berlin

Fig. 16. Displacement of epipolar curves after block adjustment by RPCs. Left images show the epipolar curve positions before and after the bundle adjustment and right images display the selected patch (identified by dashed yellow rectangles) in an enlarged image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Residuals (unit: m) of the control points for block adjustment validation. *Original* indicates the residuals before the adjustment, *modified* those after adjustment.

Area	RPCs	Mean	Max	RMSE	No. of Points
Munich	Original	1.301	2.359	1.364	31
	Modified	0.666	2.142	0.923	
Berlin	Original	0.775	1.736	0.866	32
	Modified	0.600	1.600	0.752	

However, their curvatures do not exceed more than one pixel over the whole possible range of heights in these scenes.

3.3.3. Conjugacy of epipolar curves

To investigate the conjugacy of the SAR-optical epipolar curves, two distinct points (q_1, q_2) were selected from the epipolar curve in the WorldView-2 image. From each of these points, the corresponding epipolar curves were constructed in the TerraSAR-X image for all possible heights as in the experiments before. Fig. 12 displays the corresponding epipolar curves in TerraSAR-X given by q_1 and q_2 located in the WorldView-2 image. The epipolar curve appears to pass through point p located in the SAR image. Further analysis clarifies that the differences in the column direction between the two epipolar curves passing through point p are less than one pixel, allowing the matching of the two epipolar curves (Fig. 13(a)). In addition, Fig. 14(a) shows that the gradients of the two epipolar curves change at each point, as illustrated by the column index, whereas the maximum difference is less than 0.001, i.e., 0.1%. This indicates that the epipolar curves can be assumed to be parallel. The gradient changes at each point also confirm that the epipolar curves in TerraSAR-X are not perfectly straight.

In a similar manner, the conjugacy of the epipolar curves was evaluated for the Berlin dataset. Fig. 13(b) shows that the difference between the epipolar curves is less than one pixel, so these lines can be paired. Similarly, the maximum difference between the slopes of the epipolar curves is less than 0.2%, which confirms the possibility of epipolar curve conjugacy (Fig. 14(b)).

From the above investigations and discussions, it is clear that the epipolarity constraint can be established for SAR-optical image pairs such as those from TerraSAR-X and WorldView-2 data. As expected, the epipolar curves are not perfectly straight and there are tiny differences between the epipolar curve in one image produced from points on the epipolar curve in the other image. However, analyses show that the epipolar curves can be approximated as straight lines without

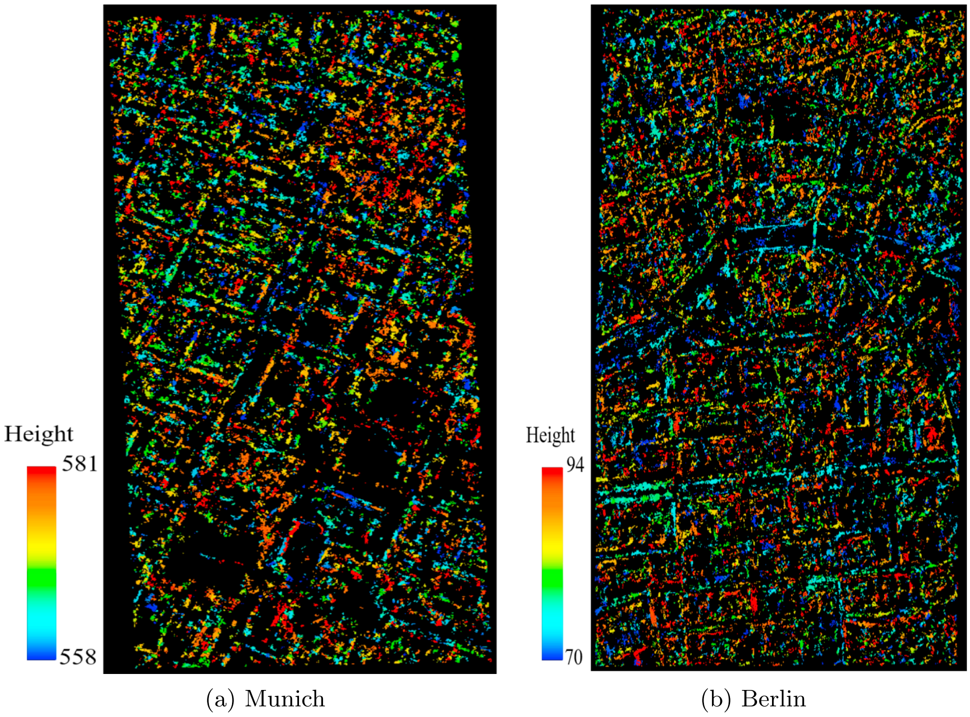


Fig. 17. Point clouds reconstructed from Munich and Berlin sub-scenes.

Table 5
Accuracy assessment of reconstructed point clouds with respect to LiDAR reference.

Area	Mean (m)			STD (m)			RMSE (m)		
	X	Y	Z	X	Y	Z	X	Y	Z
Munich	−0.003	0.025	0.080	1.285	1.350	2.652	1.285	1.351	2.653
Berlin	0.000	−0.041	0.273	1.566	1.692	3.091	1.566	1.693	3.103

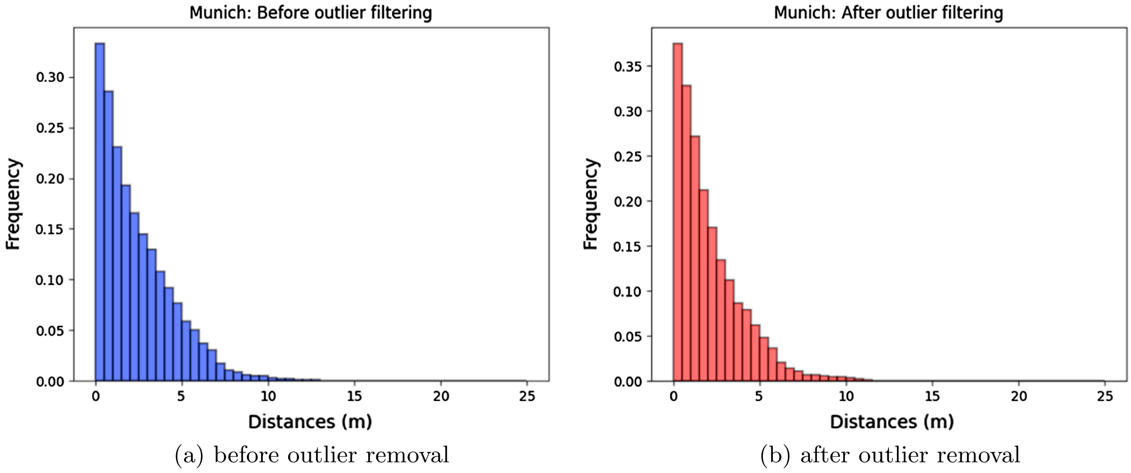
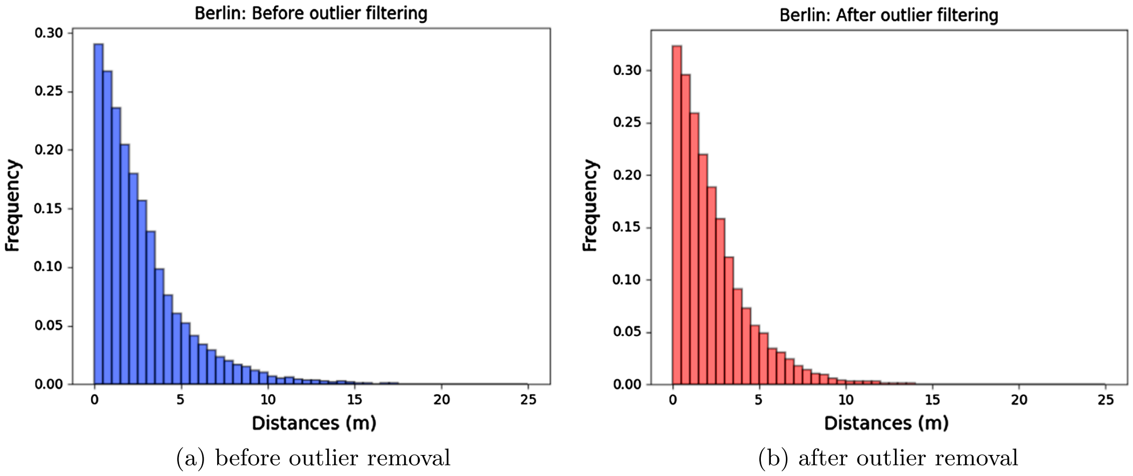


Fig. 18. Euclidian distances between reconstructed points and reference planes for Munich.

sacrificing too much, and that they can be paired together well. This means that the epipolarity constraint can be used to ease the subsequent stereogrammetric matching process.

3.4. Use of block adjustment

As discussed in Section 3.3 and mathematically proved in Section 2.2, the epipolarity constraint can be established for a SAR-optical



(a) before outlier removal

(b) after outlier removal

Fig. 19. Euclidian distances between reconstructed points and reference planes for Berlin.

Table 6
Accuracy assessment of point clouds after SRTM-based outlier removal.

Area	Point Cloud	25%-quantile	50%-quantile	75%-quantile	Mean (m)
Munich	original	0.77	1.89	3.58	2.44
	filtered	0.67	1.56	3.04	2.12
	SRTM	*0.73*	*1.64*	*3.25*	*2.21*
Berlin	original	0.89	2.01	3.67	2.75
	filtered	0.79	1.76	3.22	2.35
	SRTM	*0.86*	*1.93*	*3.63*	*2.65*

image pair. However, the positions of epipolar curves in the optical image can be placed in a more accurate position by exploiting the high geolocalization accuracy of the SAR image through a multi-sensor block adjustment. The experiments described in Section 3.3 demonstrate that, for the case of WorldView-2 imagery, the curvature of the epipolar curves does not exceed one pixel, and using only two bias terms as shifts in the column and row directions suffice to modify the position of the epipolar curves.

Implementing block adjustment using RPCs requires some conjugate points to be assigned as common tie points between the target (WorldView-2) and the reference (TerraSAR-X) images. Theoretically, just one tie point would be sufficient to estimate the bias in the least-squares adjustment based on Eq. (33) (two unknowns and two equations), but using more redundancy and incorporating more tie points allows for more accurate estimations of the bias parameter. For this experiment, eight and six tie points were selected to match the WorldView-2 images to the TerraSAR-X images in the Munich and Berlin study areas, respectively. The block adjustment equations were then established as described in Section 2.3. During the iterative least-squares adjustment, tie points with residuals exceeding a threshold were removed from the full adjustment process. Fig. 15(a) and (b) show the residuals of the full multi-sensor block adjustment for each tie point. The results demonstrate that the residuals of most points are less than one pixel in both experiments, which indicates a successful implementation of SAR-optical block adjustment. Table 3 presents the bias of the row and column components resulting from the block adjustment of WorldView-2 and TerraSAR-X image pairs for both study areas. As the SAR image was selected as the reference to which the optical imagery was aligned, the bias components for the reference SAR imagery are zero. The quality of block adjustment has been evaluated by calculating the positional errors of tie points under projection and re-projection from the SAR image to terrain and from terrain to the optical image, and vice versa. Statistical metrics such as the standard deviation (STD), median absolute deviation (MAD), and minimum (Min)/maximum (Max) errors were calculated. The results illustrate that the major bias component is in the row direction in both study areas and that the epipolar curves will mostly shift in this direction after block adjustment. Fig. 16(a) and (b) display the locations of the epipolar curves before and after adjustment. By enlarging the images, it is possible to confirm that the displacement of the epipolar curves is minimal, yet noticeable.

To verify the success of SAR-optical block adjustment, we require highly accurate GCPs, which are not available for the study areas. However, we evaluated the accuracy of block adjustment in sub-scenes of the two study areas with the assistance of available LiDAR point clouds. First, we manually found some matched points that had been measured in the SAR and optical sub-scenes, similar to the tie point selection step. Next, the measured points located in the optical imagery were projected to the terrain using the corresponding reverse rational polynomial functions ($f_o'(c, r, h)$ and $g_o'(c, r, h)$). To ensure exact back-projection, the height h of each point was extracted from the available high-resolution LiDAR point clouds of the target sub-scenes. To overcome the noise in the LiDAR data, we considered neighboring points around the selected measured point, and the final height of the target point was selected based on the mode of the heights in the considered neighborhood. The resulting ground points ($f_o'(c, r, h)$, $g_o'(c, r, h)$, h) were then back-projected to the SAR scene using the forward RPCs fitted to the SAR imagery. Finally, comparing the image coordinates of the measured points on the SAR imagery (from manual matching) with their coordinates derived by projection from the optical to the SAR imagery using RPCs and LiDAR data provides an evaluation of the SAR-optical block adjustment performance. The residuals can be calculated as:

$$\begin{aligned} dc &= f_s(f_o'(c, r, h), g_o'(c, r, h), h) - c_s^m \\ dr &= g_s(f_o'(c, r, h), g_o'(c, r, h), h) - r_s^m \end{aligned} \tag{44}$$

where (dc, dr) is the column and row difference between the measured point (c_s^m, r_s^m) located on the SAR imagery and the corresponding coordinates given by the projection from the optical to the SAR imagery using RPCs.

Table 4 presents some statistical analysis on residuals calculated according to Eq. (44) for two states: using the original WorldView-2 RPCs for the projections and using the WorldView-2 RPCs modified with respect to the TerraSAR-X SAR imagery. The results demonstrate the successful implementation of RPC-based multi-sensor block adjustment for SAR-optical image pairs. This means that the existing bias in the RPCs of optical imagery such as WorldView-2 can be modified

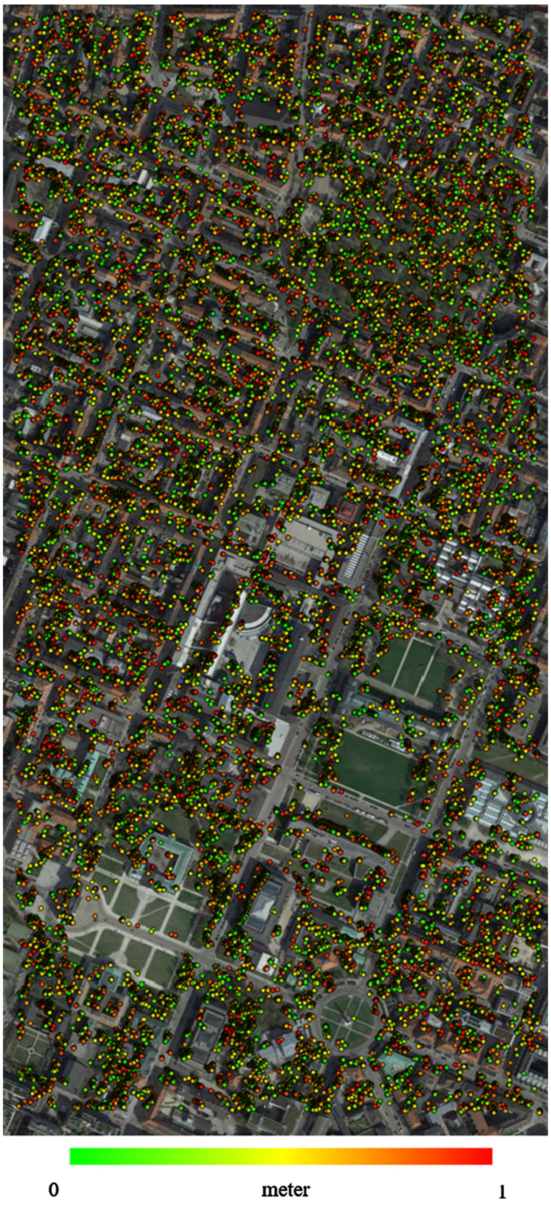


Fig. 20. Position of points with subpixel accuracy (1 m) achieved by dense matching over Munich city.

according to high-resolution SAR imagery such as TerraSAR-X to improve the absolute geolocalization accuracy of the optical imagery, and consequently the modification of epipolar curves in stereo cases.

3.5. Dense matching results

The output of dense matching by SGM is a disparity map that is calculated in the frame of the reference sensor geometry. This disparity map should be transferred from the reference sensor geometry to a terrestrial reference coordinate system such as UTM. The difference of SAR and optical observation geometries and the lack of jointly visible scene parts means that stereogrammetric 3D reconstruction leads to sparse rather than dense point clouds over urban areas. Fig. 17(a) and (b) display the reconstructed point clouds from SAR-optical sub-scenes of Munich and Berlin.

The accuracy of these sparse point clouds was compared to that of reference LiDAR point clouds with densities of 6 and 6.5 points per square meter acquired by airborne sensors over Munich and Berlin, respectively.

Different approaches can be used to assess the accuracy of point clouds. The simplest way is to calculate the Euclidean distance to the nearest-neighbor of each target point in the reference point cloud (Muja and Lowe, 2009). This strategy, however, should only be used when both point clouds are very dense. We therefore used another approach, which is based on fitting a plane to the k (here is 6) nearest neighbors of each target point in the reference point cloud (Mitra et al., 2004). The perpendicular distance from the target point to this plane is then the measured reconstruction error. To speed up the process of point cloud evaluation, an octree data structure is used for the binary partitioning of both reconstructed and reference point clouds (Schnabel and Klein, 2006). The measured distances between both point clouds are decomposed into three components that represent the accuracy of the reconstructed points along the X, Y, and Z directions. Table 5 summarizes the mean, STD, and Root Mean Square Error (RMSE) of the distances along the different axes.

In addition, histograms of the Euclidean distances between reconstructed points and k-nearest neighbors-based reference planes are depicted in Figs. 18 and 19, while the corresponding metrics are summarized in Table 6. In order to also provide an outlier-free accuracy assessment, we additionally show results corresponding to point clouds that were cleaned by removing points deviating from the SRTM model by more than 5 m.

Finally, Figs. 20 and 21 display high-accuracy points (i.e. those with a Euclidean distance of less than 1 m to the reference) achieved by SGM dense matching for TerraSAR-X- WorldView-2 image pairs in Munich and Berlin, respectively.

4. Discussion

4.1. Feasibility of SAR-optical stereogrammetry workflow

The results described in the previous section demonstrate the potential of the proposed SAR-optical stereogrammetry framework. Our analyses show that all primary steps involved in SAR-optical stereogrammetry, such as RPC fitting, epipolar-curve generation, and multi-sensor block adjustment, can be successfully implemented for VHR SAR-optical image pairs. In addition to the mathematical proof of the existence of an epipolarity constraint for arbitrary SAR-optical image pairs in Section 2.2, the experimental results have illustrated the validity of establishing an epipolarity constraint by showing that SAR-optical epipolar curves are approximately straight. Using RPCs for both sensor types paves the way for the implementation of stereogrammetry. As a result, estimating the RPCs for SAR imagery is a prerequisite for SAR-optical stereogrammetry. The RPCs delivered with optical imagery must be improved with respect to the SAR sensor geometry using RPC-based multi-sensor block adjustment. The block adjustment aligns pairs of SAR and optical images and improves the absolute geopositioning accuracy of the optical imagery. This ensures that the epipolar curves pass through the correct positions of conjugate points. Applying a dense matching algorithm such as SGM then produces a disparity map.

4.2. Potential and limitations of SAR-optical stereogrammetry

As discussed in Section 3.5, the dense matching of TerraSAR-X/WorldView-2 imagery produces a sparse point cloud over each of the urban study areas. However, the resulting point clouds are affected by a significant amount of noise because of the difficult radiometric and geometric relationships between the SAR and the optical images. Hence, the SGM algorithm struggles to find the exact conjugate points. On the one hand, this is related to the similarity measures employed in

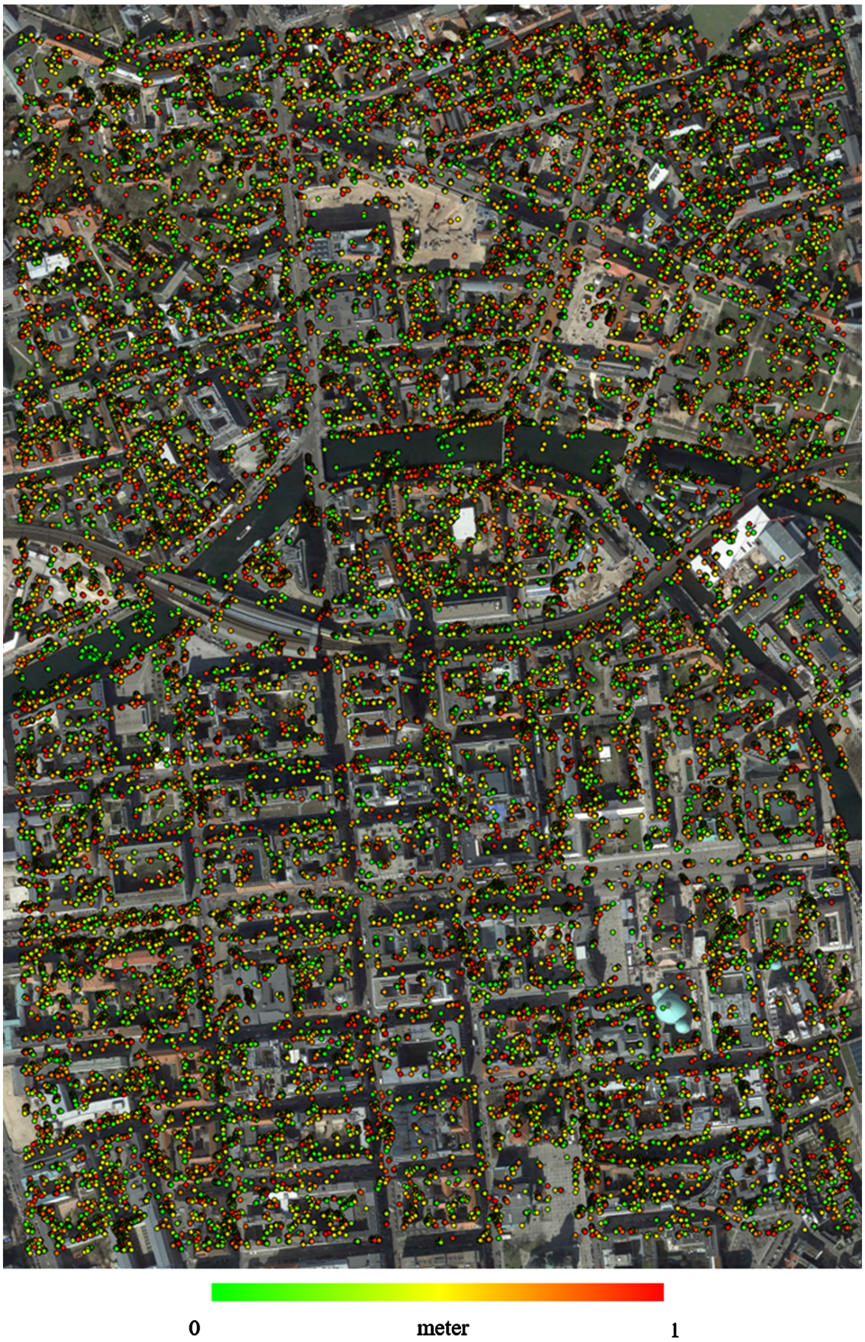


Fig. 21. Position of points with subpixel accuracy (1 m) achieved by dense matching over Berlin city.

this prototypical study. The influence of similarity measures on the height accuracy of the Munich point cloud is shown in Fig. 22.

The RMSE of the estimated heights decreases when Census and MI are combined and used as a weighted sum cost function, although the number of outliers increases. Identifying the optimum weighting to balance the percentage of outliers against the height accuracy is impractical, because the output disparity maps are rather sparse; a visualization would not be helpful for this task. In stereogrammetric 3D reconstruction using optical image pairs, visualizing the disparity map enables the weight value to be tuned so as to preserve the edges and sharpness of building footprints, whereas in the SAR-optical case, there is no perfectly dense disparity map. Using Census alone produces points with higher accuracy than the MI-only results, but a higher percentage of outliers. In general, a similarity measure specifically designed for SAR-optical matching is required.

On the other hand, the reconstruction suffers from the fact that the SGM search strategy is designed for relatively simple isotropic geometric distortions and was not adapted to the peculiarities of SAR-optical matching yet. Therefore, the differences in the imaging geometries of SAR and optical sensors in terms of their off-nadir and horizontal

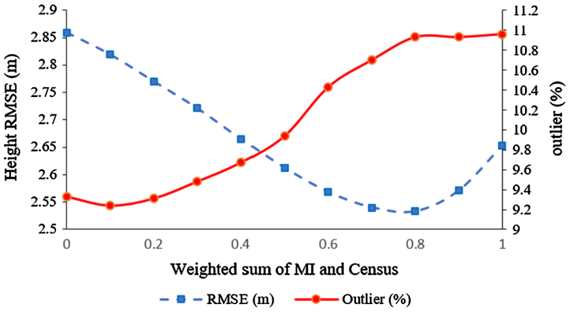


Fig. 22. Performance of MI, Census, and weighted sum of both measures as cost functions in SGM.

viewing angles can further decrease the matching accuracy. For example it is known from previous research that the optimal geometrical condition for SAR-optical stereogrammetry would be an image pair acquired with similar viewing geometries (Qiu et al., 2018). This, however, would make the geometrically induced dissimilarities in the images even larger and render the matching more complicated. If both sensors were at the same position, and thus would share the same viewing angle, the intersection geometry would be perfect (Qiu et al., 2018). However, due to the different imaging geometries, elevated objects would appear to collapse away from the sensor in the optical image, while they would appear to collapse towards the sensor in the SAR image. Thus, the choice of a good stereo geometry will always need to be a trade-of between image similarity and favorable intersection angle in the SAR-optical case.

Last, but not least, many points cannot be sensed by a nadir-looking optical sensor but are well observed by a side-looking SAR sensor, such as points located on building facades. As has been shown before, the joint visibility between SAR and optical VHR images of urban scenes can be as low as about 50% (Hughes et al., 2018). In the present study, the situation was most complicated in the Berlin case, because of differences in both the horizontal viewing directions and the off-nadir angles. The horizontal viewing direction of the WorldView-2 sensor was approximately north-south, whereas that of TerraSAR-X was east-west. This affected the visibility of common points between the two images during 3D reconstruction negatively. Consequently, most of the reconstructed points are located on the flat areas or outlines of buildings that are observed by both sensors (see Figs. 20 and 21).

Finally, some differences in the image pairs may be caused by the interval between the acquisition times of the WorldView-2 and TerraSAR-X data (5 and 3 years for Munich and Berlin, respectively). This can cause the matching process to fail in problematic areas, thus affecting the quality and density of the disparity maps.

In spite of the differences in sensor geometries, acquisition times, and illumination conditions between the two SAR-optical image pairs, the quantitative analysis demonstrated in Section 3.5 shows that 25% of all points are reconstructed with clear sub-pixel accuracy, while the median accuracy lies at about 1.5–2 m. The experiments also show that the results can be further improved by filtering outliers from the reconstructed point clouds. In this study, we employed the globally available SRTM DEM as prior knowledge for outlier removal. As Table 6 shows, discarding points with a height difference to SRTM greater than 5 m improves the results significantly.

Of course, this simple filtering strategy will probably also remove some accurate points that just deviate a lot from the SRTM DEM (e.g. newly built skyscrapers). In conclusion, a more sophisticated algorithm should be developed for removing noise and outliers from derived point clouds in the future. Nevertheless it can be confirmed that the SAR-optical stereo results have the potential to provide both higher accuracy and higher point density than the SRTM data, making SAR-optical stereogrammetry another possible means for 3D reconstruction in remote sensing.

5. Conclusion

In this study, we investigated the possibility of stereogrammetric 3D reconstruction from VHR SAR-optical image pairs by developing a full 3D reconstruction framework based on the classic photogrammetric workflow. First, we analyzed all prerequisites for this task. The main requirement for SAR-optical stereogrammetry is to establish an epipolarity constraint to reduce the search space of the matching process. We mathematically proved that the epipolarity constraint can be established for SAR-optical image pairs. Furthermore, experimental analysis demonstrated that the epipolarity constraint can be employed for SAR-optical image pairs such as those from TerraSAR-X/WorldView-2, and showed that the epipolar curves are sufficiently straight. Because of the limited accuracy of the RPCs delivered with optical data, the relative orientation between both images can be improved with respect to the more accurate SAR orientation parameters using multi-sensor block adjustment. This shifts the epipolar curves toward their correct positions. An SGM-based dense matching algorithm was implemented using the MI and Census similarity measures, as well as their weighted sum. The outputs were sparse point clouds with a median accuracy of about 1.5 to 2 m and the 25%-quantile of best points well in the sub-pixel accuracy domain. Finally, SRTM data were used to remove outliers from the point clouds. This improved the accuracy of the point clouds further. Overall, this study has demonstrated that a 3D reconstruction framework can be designed and implemented for SAR-optical image pairs over urban areas. Future work will have to focus on the development of similarity metrics specific to the multi-sensor matching problem, and on an adaption of the semi-global search strategy that accounts for the anisotropic geometric distortions between SAR and optical images.

Acknowledgments

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, the Bavarian Surveying Administration for the LiDAR reference data of Munich, and Land Berlin (EU EFRE project) for the LiDAR reference data of Berlin.

This work is jointly supported by the German Research Foundation (DFG) under grant SCHM 3322/1-1, the Helmholtz Association under the framework of the Young Investigators Group SiPEO (VH-NG-1018), and the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. ERC-2016-StG-714087, Acronym: So2Sat).

References

Bagheri, H., Schmitt, M., d'Angelo, P., Zhu, X.X., 2018. Exploring the applicability of semi-global matching for SAR-optical stereogrammetry of urban scenes. ISPRS Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci. 42 (2), 43–48.

Bloom, A.L., Fielding, E.J., Fu, X.-Y., 1988. A demonstration of stereophotogrammetry with combined SIR-B and Landsat TM images. Int. J. Remote Sens. 9, 1023–1038.

Brown, M.Z., Burschka, D., Hager, G.D., 2003. Advances in computational stereo. IEEE Trans. Pattern Anal. Mach. Intell. 25, 993–1008.

Cho, W., Schenk, T., Madani, M., 1993. Resampling digital imagery to epipolar geometry. Int. Arch. Photogram. Remote Sens. 29, 404.

Curlander, J.C., 1982. Location of spaceborne SAR imagery. IEEE Trans. Geosci. Remote Sens. 22, 106–112.

Curlander, J.C., McDonough, R.N., 1991. Synthetic Aperture Radar, vol. 396 John Wiley & Sons, New York, NY, USA.

d'Angelo, P., Reinartz, P., 2012. DSM based orientation of large stereo satellite image blocks. ISPRS Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci. 39 (B1), 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).

Egnal, G., Wildes, R.P., 2002. Detecting binocular half-occlusions: empirical comparisons

of five approaches. IEEE Trans. Pattern Anal. Mach. Intell. 24, 1127–1133.
Eineder, M., Minet, C., Steigenberger, P., Cong, X., Fritz, T., 2011. Imaging geodesy–toward centimeter-level ranging accuracy with TerraSAR-X. IEEE Trans. Geosci. Remote Sens. 49, 661–671.
Fraser, C.S., Hanley, H.B., 2005. Bias-compensated RPCs for sensor orientation of high-resolution satellite imagery. Photogram. Eng. Remote Sens. 71, 909–915.
Fraser, C., Dial, G., Grodecki, J., 2006. Sensor orientation via RPCs. ISPRS J. Photogram. Remote Sens. 60, 182–194.
Grodecki, J., Dial, G., 2003. Block adjustment of high-resolution satellite images described by rational polynomials. Photogram. Eng. Remote Sens. 69, 59–68.
Grodecki, J., Dial, G., Lutes, J., 2004. Mathematical model for 3D feature extraction from multiple satellite images described by RPCs. In: ASPRS Annual Conference Proceedings, Denver, Colorado.
Gupta, R., Hartley, R.I., 1997. Linear pushbroom cameras. IEEE Trans. Pattern Anal. Mach. Intell. 19, 963–975.
Gutjahr, K., Perko, R., Raggam, J., Schardt, M., 2014. The epipolarity constraint in stereo-radargrammetric DEM generation. IEEE Trans. Geosci. Remote Sens. 52, 5014–5022.
Hartley, R.I., Zisserman, A., 2004. Multiple View Geometry in Computer Vision, second ed. Cambridge University Press ISBN: 0521540518.
Hassaballah, M., Abdelmgeid, A.A., Alshazly, H.A., 2016. Image features detection, description and matching. In: Awad, A.I., Hassaballah, M. (Eds.), Image Feature Detectors and Descriptors: Foundations and Applications. Springer International Publishing, pp. 11–45.
Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. 30, 328–341.
Hughes, L.H., Auer, S., Schmitt, M., 2018. Investigation of joint visibility between SAR and optical images of urban environments. ISPRS Ann. Photogram. Remote Sens. Spatial Inform. Sci. 4, 129–136.
Kim, T., 2000. A study on the epipolarity of linear pushbroom images. Photogram. Eng. Remote Sens. 66, 961–966.
Kratky, V., 1989. Rigorous photogrammetric processing of SPOT images at CCM Canada. ISPRS J. Photogram. Remote Sens. 44, 53–71.
Li, D., Zhang, Y., 2013. A rigorous SAR epipolar geometry modeling and application to 3D target reconstruction. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 6, 2316–2323.
Li, R., Zhou, F., Niu, X., Di, K., 2007. Integration of Ikonos and QuickBird imagery for geopositioning accuracy analysis. Photogram. Eng. Remote Sens. 73, 1067.
Merkle, N., Luo, W., Auer, S., Müller, R., Urtasun, R., 2017. Exploiting deep matching and SAR data for the geo-localization accuracy improvement of optical satellite images. Remote Sens. 9, 586.
Mitra, N.J., Nguyen, A., Guibas, L., 2004. Estimating surface normals in noisy point cloud data. Int. J. Comput. Geom. Appl. 14, 261–276.
Morgan, M., Kim, K., Jeong, S., Habib, A., 2004. Epipolar geometry of linear array scanners moving with constant velocity and constant attitude. ISPRS Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci. 35, 508–513.
Morgan, M.F., El-Sheimy, N., Saalfeld, A., 2004. Epipolar resampling of linear array scanner scenes (PhD dissertation). University of Calgary, Department of Geomatics Engineering.
Muja, M., Lowe, D.G., 2009. Fast approximate nearest neighbors with automatic algorithm configuration. In: VISAPP International Conference on Computer Vision Theory and Applications, pp. 331–340.
Oh, J., Lee, W.H., Toth, C.K., Grejner-Brzezinska, D.A., Lee, C., 2010. A piecewise approach to epipolar resampling of pushbroom satellite images based on RPC. Photogram. Eng. Remote Sens. 76, 1353–1363.
Orun, A.B., 1994. A modified bundle adjustment software for SPOT imagery and photography: tradeoff. Photogram. Eng. Remote Sens. 60, 1431–1437.
Perko, R., Raggam, H., Gutjahr, K., Schardt, M., 2011. Using worldwide available TerraSAR-X data to calibrate the geo-location accuracy of optical sensors. In: 2011 IEEE International Geoscience and Remote Sensing Symposium, pp. 2551–2554.
Qiu, C., Schmitt, M., Zhu, X.X., 2018. Towards automatic SAR-optical stereogrammetry over urban areas using very high resolution imagery. ISPRS J. Photogram. Remote Sens. 138, 218–231.
Raggam, J., Almer, A., Strobl, D., 1994. A combination of SAR and optical line scanner imagery for stereoscopic extraction of 3D data. ISPRS J. Photogram. Remote Sens. 49, 11–21.
Scharstein, D., Szeliski, R., Zabih, R., 2001. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001), pp. 131–140.
Schmitt, M., Zhu, X.X., 2016. Data fusion and remote sensing: an ever-growing relationship. IEEE Geosci. Remote Sens. Mag. 4, 6–23.
Schnabel, R., Klein, R., 2006. Octree-based point-cloud compression. In: Proceedings of the 3rd Eurographics/IEEE VGTC Conference on Point-Based Graphics, SPBG'06. Eurographics Association, Aire-la-Ville, Switzerland, pp. 111–121.
Suri, S., Reinartz, P., 2010. Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas. IEEE Trans. Geosci. Remote Sens. 48, 939–949.
Tao, C., Hu, Y., 2001. Use of the rational function model for image rectification. Can. J. Remote Sens. 27, 593–602.
Tao, C.V., Hu, Y., 2001. A comprehensive study of the rational function model for photogrammetric processing. Photogram. Eng. Remote Sens. 67, 1347–1358.
Tao, C.V., Hu, Y., 2002. 3D reconstruction methods based on the rational function model. Photogram. Eng. Remote Sens. 68, 705–714.
Tao, C.V., Hu, Y., Jiang, W., 2004. Photogrammetric exploitation of Ikonos imagery for mapping applications. Int. J. Remote Sens. 25, 2833–2853.
Tikhonov, A., Arsenin, V.Y., 1977. Methods for Solving Ill-posed Problems. John Wiley and Sons, Inc.
Tong, X., Liu, S., Weng, Q., 2010. Bias-corrected rational polynomial coefficients for high accuracy geo-positioning of QuickBird stereo imagery. ISPRS J. Photogram. Remote Sens. 65, 218–226.
Toutin, T., 2006. Comparison of 3D physical and empirical models for generating DSMs from stereo HR images. Photogram. Eng. Remote Sens. 72, 597–604.
USGS, 2000. Shuttle Radar Topography Mission (SRTM) void filled. < https://lta.cr.usgs.gov/SRTMVF > (accessed 09.17).
Wegner, J.D., Ziehn, J.R., Soergel, U., 2014. Combining high-resolution optical and InSAR features for height estimation of buildings with flat roofs. IEEE Trans. Geosci. Remote Sens. 52, 5840–5854.
Zhang, L., He, X., Balz, T., Wei, X., Liao, M., 2011. Rational function modeling for spaceborne SAR datasets. ISPRS J. Photogram. Remote Sens. 66, 133–145.
Zhu, K., d'Angelo, P., Butenuth, M., 2011. A performance study on different stereo matching costs using airborne image sequences and satellite images. In: Stilla, U., Rottensteiner, F., Mayer, H., Jutzi, B., Butenuth, M. (Eds.), Photogrammetric Image Analysis. Springer, Berlin, Heidelberg, pp. 159–170.

A.4 H. Bagheri, M. Schmitt, and X. X. Zhu. Fusion of multi-sensor-derived heights and OSM-derived building footprints for urban 3D reconstruction. In: ISPRS International Journal of Geo-Information 8.4 (2019).

International Journal of
Geo-Information

Article

Fusion of Multi-Sensor-Derived Heights and OSM-Derived Building Footprints for Urban 3D Reconstruction

Hossein Bagheri [1], **Michael Schmitt** [1] **and Xiaoxiang Zhu** [1,2,*]

[1] Signal Processing in Earth Observation, Technical University of Munich, 80333 Munich, Germany; hossein.bagheri@tum.de (H.B.); m.schmitt@tum.de (M.S.)
[2] Remote Sensing Technology Institute, German Aerospace Center, 82234 Wessling, Germany
[*] Correspondence: xiaoxiang.zhu@dlr.de; Tel.: +49-89-289-22659

Received: 28 February 2019; Accepted: 9 April 2019; Published: 18 April 2019

Abstract: So-called prismatic 3D building models, following the level-of-detail (LOD) 1 of the OGC City Geography Markup Language (CityGML) standard, are usually generated automatically by combining building footprints with height values. Typically, high-resolution digital elevation models (DEMs) or dense LiDAR point clouds are used to generate these building models. However, high-resolution LiDAR data are usually not available with extensive coverage, whereas globally available DEM data are often not detailed and accurate enough to provide sufficient input to the modeling of individual buildings. Therefore, this paper investigates the possibility of generating LOD1 building models from both volunteered geographic information (VGI) in the form of OpenStreetMap data and remote sensing-derived geodata improved by multi-sensor and multi-modal DEM fusion techniques or produced by synthetic aperture radar (SAR)-optical stereogrammetry. The results of this study show several things: First, it can be seen that the height information resulting from data fusion is of higher quality than the original data sources. Secondly, the study confirms that simple, prismatic building models can be reconstructed by combining OpenStreetMap building footprints and easily accessible, remote sensing-derived geodata, indicating the potential of application on extensive areas. The building models were created under the assumption of flat terrain at a constant height, which is valid in the selected study area.

Keywords: 3D building reconstruction; building model; OpenStreetMap (OSM); building foot prints; multi-sensor fusion; digital elevation models (DEM); LOD1; SAR-optical stereogrammetry

1. Introduction

One particular interest in remote sensing is the 3D reconstruction of urban areas for diverse applications such as 3D city modeling, urban, and crisis management, etc. Buildings belong to the most important objects in urban scenes and are modeled for diverse applications such as simulation of air pollution, estimating energy consumption, detecting urban heat islands, and many others [1]. There are different levels of building modeling which have been described under the standard of the OGC City Geography Markup Language (CityGML). These are summarized in [2].

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 describes building models as block models with flat roof structure and provides the coarsest volumetric representation of buildings [3]. Thus, LOD1 models are frequently produced by extruding a building footprint to a height provided by separate sources [4]. The next level is LOD2, which represents building shapes with more details. Therefore, this type of building modeling demands high-resolution data in comparison to the first level.

Comprehensive technical information about variants of the LOD of a 3D building model can be found in [5]. In many cases, the building height information can be provided by versatile remote sensing data sources such as airborne laser scanning [6], high-resolution optical stereo imagery [7], or DEMs produced by synthetic aperture radar (SAR) interferometry [8]. Other sources for LOD modelling are described in [9].

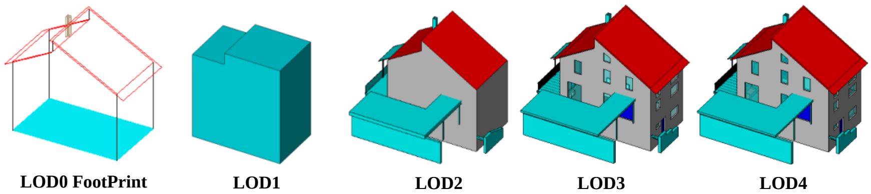


Figure 1. Different levels of detail of building models according to OGC City Geography Markup Language (CityGML) 2.0 [10].

A special interest lies in automatically generating building models for extensive areas at LOD1 level. While height information provided by airborne LiDAR data leads to highly accurate LOD1 representations of buildings [11,12], it is computationally expensive to produce models that cover wide areas. In addition, expensive LiDAR data are often not available for extensive areas. On the other hand, several investigations illustrate the possibility of using other remote sensing data types for 3D building reconstruction for that purpose [13,14]. As an example, the possibility of LOD1 3D building model generation from Cartosat-1 and Ikonos DEMs has been investigated in [15]. In another study, Marconcini et al. proposed a method for building height estimation from TanDEM-X data [16]. Using open DEMs such as SRTM for 3D reconstruction has been evaluated in different studies [17–19]. They concluded that SRTM elevation data can be used for recognizing tall buildings. In a recent investigation, Misra et al. compared different global height data sources such as SRTM, ASTER, AW3D, as well as TanDEM-X for digital building height model generation [20].

The main objective of this paper is to investigate the possibility of LOD1-based 3D building modeling from different remote sensing data sources which can be efficiently applied to wide areas. Regarding that each remote sensing source provided by a sensor with specific properties, using multi-sensor data fusion techniques can ultimately provide high quality geodata for 3D reconstruction by instructively integrating the sensors' properties and mitigating their drawbacks [21]. For that purpose, height information is extracted from different sources: medium-resolution DEMs derived from optical imagery such as the Cartosat-1 DEM, and interferometric DEMs generated from bistatic TanDEM-X acquisitions. Due to the limitations and specific properties of those DEMs, state-of-the art DEM fusion techniques are used for improving the height accuracy. More details of those techniques and the logic behind the fusion are explained in the respective sections.

In another experiment, the potential of using heights from SAR-optical stereogrametry for 3D building reconstruction is investigated. Regarding the growing archive of very high-resolution SAR and optical imagery, developing a framework that takes advantages of both SAR and optical imagery can provide a great opportunity to produce 3D spatial information over urban areas. Besides the globally available DEMs derived from optical and SAR remote sensing, this information can also potentially be employed for producing 3D building models at LOD1 level.

Besides height data, building outlines are needed for LOD1 modelling, since the aforementioned height sources are not detailed enough to reliably determine accurate building outlines. We therefore use OpenStreetMap as a form of volunteered geographic information (VGI) that is available with global coverage as well. In this paper, we evaluate the potential of 3D building reconstruction from both building footprints provided by OSM and heights derived by multi-sensor remote sensing data fusion. Since the study area in this research is flat, we consider a constant height for ground and finally generate a building model with this assumption.

In Section 2, different fusion techniques used for height derivation over urban areas are summarized. It includes three fusion experiments: TanDEM-X and Cartosat-1 DEM fusion (Section 2.1), multiple TanDEM-X raw DEM fusion (Section 2.2), and SAR-optical stereogrammetry for 3D urban reconstruction (Section 2.3). After that, a simple procedure for LOD1 building model reconstruction from the multi-sensor-fusion-derived heights and OSM building footprints is presented in Section 3. The properties of the applied data and the study area are described in Section 4, including a summary of the benefits of multi-sensor DEM fusion and SAR-optical stereogrammetry. The outputs and results of LOD1 building model reconstruction using both VGI and different remote-sensing-derived geodata are provided in Section 5. Finally, the potential of LOD1 3D reconstruction using the mentioned data sources, as well as challenges and open issues, are discussed in Section 6.

2. Multi-Sensor Data fusion for Height Generation over Urban Scenes

In this paper, elevation data are derived from different sensor types for 3D building reconstruction. As mentioned earlier, those data sources can be categorized as digital elevation models derived from optical or SAR imagery and also as point clouds reconstructed from SAR-optical image pairs through stereogrammetry. The main idea is to apply data fusion techniques to finally produce more accurate height information. In the following sections, more details of applied fusion techniques will be presented.

2.1. TanDEM-X and Cartosat-1 DEM Fusion in Urban Areas

Cartosat-1 is an Indian satellite equipped with optical sensors for stereo imagery acquisitions. The Cartosat-1 sensor with resolution of 2.5 m and partially large swath width of 30 km makes the acquired stereo images perfect for producing high-resolution DEMs with a wide coverage [22]. However, the main defect of this sensor is the poor absolute localization accuracy [23]. In parallel, the TanDEM-X mission is a recent endeavour for producing a global DEM through an interferometric SAR processing chain. Evaluation with respect to LiDAR reference data illustrates that the TanDEM-X DEM has a better absolute accuracy than the Cartosat-1 DEM, while its precision drops out in urban areas because of intrinsic properties of InSAR-based height construction [24]. Figure 2b shows the performance of both DEMs in a subset selected for height precision evaluation over an urban scene. As displayed in Figure 2b, the overall precision of the Cartosat-1 DEM is better than the overall precision of the TanDEM-X DEM.

Regarding the drawbacks of both DEMs, data fusion is used to finally reach a high quality DEM. In more detail, first the absolute accuracy of Cartosat-1 is increased to the level of absolute accuracy of the TanDEM-X DEM by vertical alignment. Next, both DEMs can be integrated using a sophisticated approach presented in our previous research [25]. The fusion method is developed for multi-sensor DEM fusion with the support of neural-network-predicted fusion weights. For this task, appropriate spatial features are extracted from both target DEMs as well as respective height residuals from some training subsets. The height residuals are calculated respective to available LiDAR over training data. After that, a refinement process is carried out to explore numerical feature-error relations between each type of extracted features and height residuals. Then, the refined feature-error relations are input into fully-connected neural networks to predict a weight map for each DEM. The predicted weight maps can be applied for weighted averaging-based fusion of the input Cartosat-1 and TanDEM-X DEMs. Figure 3 displays the designed pipeline for ANN-based fusion of TanDEM-X and Cartosat-1 DEMs.

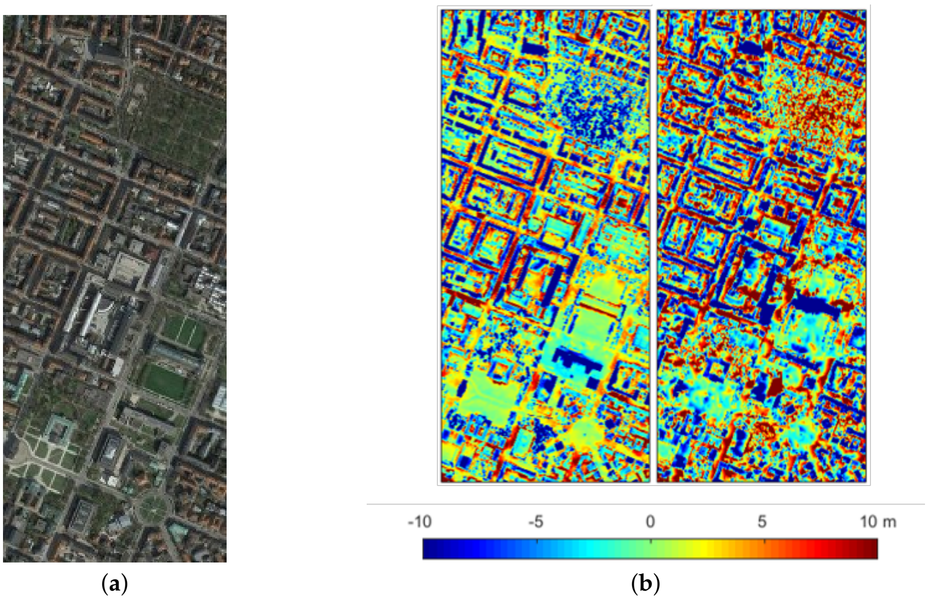


Figure 2. (**a**) Study subset selected over Munich, (**b**) Precision of the Cartosat-1 (left) and TanDEM-X (right) digital elevation models (DEMs) over an exemplary urban subset respective to high-resolution LiDAR data. Both DEMs were assessed with respect to a co-aligned LiDAR DEM.

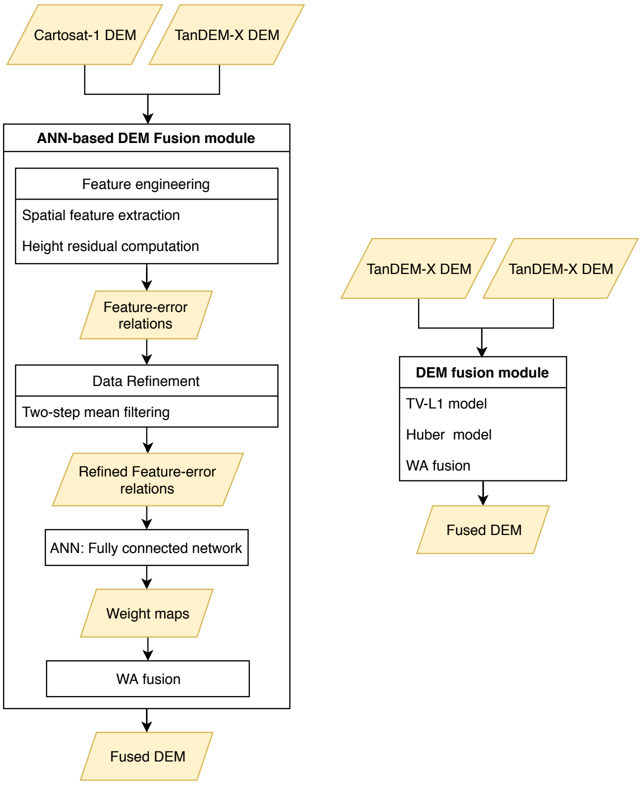


Figure 3. Different DEM fusion modules for improving the TanDEM-X quality. Left: The proposed pipeline for TanDEM-X and Cartosat-1 DEM fusion, Right: Process of multi-modal TanDEM-X DEM fusion.

2.2. TanDEM-X Raw DEM Fusion over Urban Areas

As mentioned earlier, another possibility to gather reliable height information is to fuse multi-modal TanDEM-X raw DEMs. The standard TanDEM-X DEM is the output of a processing chain consisting of interferometry, phase unwrapping (PU), data calibration, DEM block adjustment, and raw DEM mosaicking [26]. In the mosaicking step, raw DEMs are fused to reach the target accuracy. The fusion method is weighted averaging using weights derived from a height error map produced during the interferometry process. Evaluation demonstrates that weighted averaging does not perform well in urban areas. We proposed to use a more sophisticated fusion approach for fusing TanDEM-X raw DEMs in [27]. For this, we used variational models like TV-L_1 and Huber models and finally produced a high quality DEM over urban areas in comparison to weighted averaging. In this paper, we also apply TV-L_1 and Huber models for fusion of TanDEM-X raw DEMs over the study urban subset to improve height accuracy for 3D building reconstruction. A comparison between the multi-modal TanDEM-X DEM fusion process and the multi-sensor ANN-based fusion is depicted in Figure 3.

2.3. Heights from SAR-Optical Stereogrammetry

In the literature, a few papers can be found that deal with the combination of SAR and optical imagery for the 3D reconstruction of urban objects, e.g., [28]. In this research, we focus on the potential of 3D building reconstruction from very high-resolution SAR-optical image pairs such as TerraSAR-X/WorldView-2 through a dense matching process as a form of cooperative data fusion [21].

A full framework for stereogrammetric 3D reconstruction from SAR-optical image pairs was presented in our previous work [29] is displayed in Figure 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.

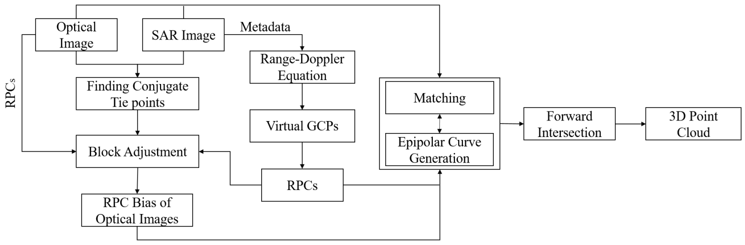


Figure 4. Framework for 3D reconstruction from synthetic aperture radar (SAR)-optical image pairs [29].

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 [30]. 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 more robust and lighter than other typical dense matching methods [30], which can be ptentially applied for SAR-optical stereogrammetry. According to [31], pixel-wise Mutual information (MI), and Census are more appropriate for difficult illumination relationships than, e.g., normalized cross-correlation (NCC).

3. LOD1 Building Model Generation

The heights output by the different fusion approaches are then used for 3D building modeling and finally prismatic model generation. Due to the medium resolution of the input DEMs, only LOD1 models can be reconstructed from those heights; also the resolutions of the DEMs are not sufficient for detecting building outlines. As shown in Section 4.3, the point cloud resulting from SAR-optical stereogrammetry is partially sparse and consequently building outlines can not be recognized. One popular option is to exploit the building footprints layer provided by OpenStreetMap (OSM). Then, the heights of building outlines can be derived from either those fused DEMs or the point cloud achieved by SAR-optical stereogrammetry. Technically, this can be realized in two steps. The first step is to classify heights to those located inside and outside building outlines. Then, only points that are within building outlines are kept while the remaining points are discarded. After that, for each remaining height, the ID of the corresponding building (in which the height is located) is assigned. It facilitates the process of joining building footprints layer to heights.

There are several elevation references that should be considered for estimating the building height within its outline [32]. These references are displayed in Figure 5. Three-dimensional reconstruction based on those levels can be realized by using high-resolution data such as LiDAR point clouds along with precise cadastral maps. Specifying those levels in medium resolution remote-sensing-derived heights, however, is not possible. Therefore, for LOD1 3D building reconstruction using medium resolution data such as those applied in this paper, we will only use median or mean of heights inside a building outline. The main advantage of median is its robustness against outliers in comparison to the mean measure. Thus, we propose that LOD1 models can be produced by modeling each building as a coarse volumetric representation using its outline and the median-based allocated height.

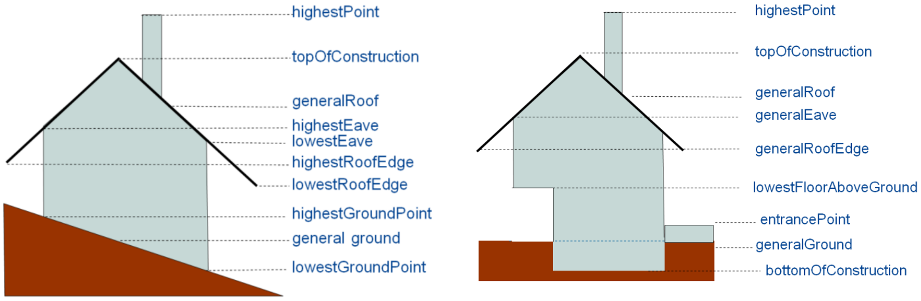


Figure 5. Examples of elevation references for different kinds of building [32].

Furthermore, for LOD1 reconstruction, we will consider two scenarios. The first one is to model buildings based on the original footprint layers provided by OSM. The second is to update these building outlines in a pre-processing step. This updating has proved to be helpful, because of OSM building footprints often consist of several intra-blocks with different heights. As displayed in Figure 1, a building consisting of two blocks, each with different height level, may appear as an integrated building outline in OSM and thus, only one height value could be assigned for it in a simple LOD1 reconstruction process, while the outline should actually be split into two separate outlines. The result will be that the heights that actually lie in two separate clusters will erroneously be substituted by their median value located somewhere in the middle. While this ultimately leads to a significant height bias, modifying the outlines appropriately optimizes the final reconstruction. In this paper, this building modification is performed semi automatically: The candidate outlines are detected by clustering heights. The number of clusters determines the number of height levels and implies potential separate building blocks. Then, this is verified by visual comparison with open satellite imagery such as provided by Google Earth. Finally, the individual, newly separated building blocks are reconstructed by assigning separate median height values.

In addition to that, horizontal displacements of OSMs' building footprints respective to highly accurate data such as LiDAR can also lead to a height bias. This phenomenon leads to an inclusion of non-building points to building outlines. Due to significant height differences between non-building and building points, the final height estimations are affected by an underestimation bias. To mitigate this effect, we use a buffer from the building outline inwards to make sure only building points are selected.

4. Test Data

In this paper, as explained in Section 2, the heights for 3D building reconstruction are provided by different sources. For the experiments, a study scene located in Munich, Germany, was selected because of the availability of high-quality LiDAR reference data. Figure 2a displays the considered study urban subset. The characteristics of the different input datasets used in the experiments are listed in following.

- *Cartosat-1 DEM*: The Cartosat-1 DEM used in this study is produced from stacks of images acquired over the Munich area based on the pipeline described in [33]. The main characteristics of the Cartosat-1 DEM are expressed in Table 1.

Table 1. Properties of Cartosat-1 tile. For more information about BKG orthophotos, please refer to [34].

Cartosat-1 DEM	
Stereoscopic angle	31°
Max number of rays	11
Min number of rays	2
Horizontal reference	BKG orthophotos
Vertical reference	SRTM DEM
Pixel spacing	5 m
Mean height error (1σ)	2–3 m

- *TanDEM-X raw DEMs*: In this study two tiles of TanDEM raw DEM acquired over Munich city are used. The properties of those tiles are represented in Table 2.

Table 2. Properties of the nominal TanDEM-X raw digital elevation models (DEMs) tiles for the Munich area.

TanDEM-X Raws DEMs: Munich Area		
Acquisition Id	**1023491**	**1145180**
Acquisition mode	Stripmap	Stripmap
Center incidence angle	38.25°	37.03°
Equator crossing direction	Ascending	Ascending
Look direction	Right	Right
Polarization	HH	HH
Height of ambiguity	45.81 m	53.21 m
Pixel spacing	0.2 arcsec	0.2 arcsec
HEM mean	1.33 m	1.58 m

- *TerraSAR-X and WordView-2 images*: For the experiment based on heights retrieved by SAR-optical stereogrammetry, a high-resolution TerraSAR-X/WorldView-2 image pair, acquired over the Munich test scene, is used. For the pre-processing, first, the SAR image was filtered by a non-local filter to reduce the speckle [35]. After that, they were resampled to 1 m × 1 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 Figure 6. The specifications of the TerraSAR-X and WorldView-2 images are provided in Table 3.

Table 3. Specifications of the TerraSAR-X and WorldView-2 images.

Sensor	Acquisition Mode	Off-Nadir Angle (°)	Ground Pixel Spacing (m)	Acquisition Date
TerraSAR-X	Spotlight	22.99	0.85 × 0.45	03.2015
WorldView-2	Panchromatic	5.20	0.50 × 0.50	07.2010

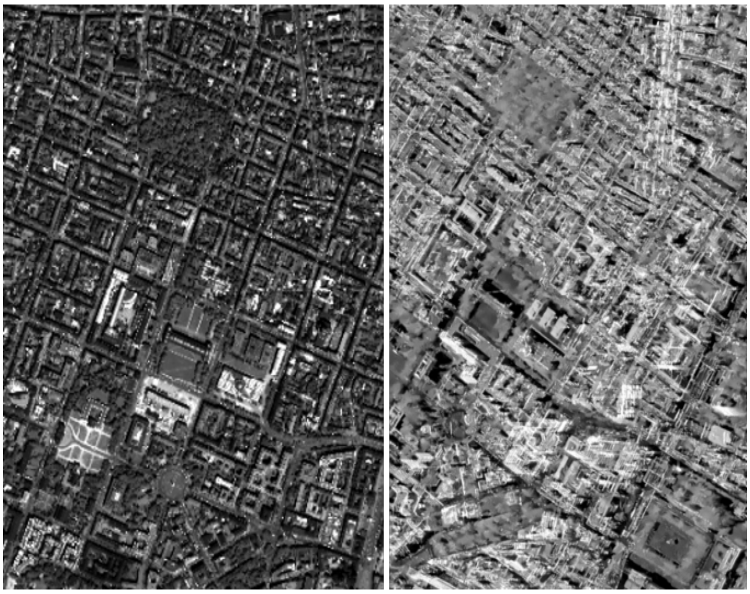


Figure 6. Display of SAR-optical sub-scenes extracted from Munich study areas (the left-hand image is from WorldView-2, the right-hand image is from TerraSAR-X).

- *LiDAR point cloud*: High-resolution airborne LiDAR data serves for performance assessment and accuracy evaluation of 3D building reconstruction resulting from different height information sources. It is also used for measuring accuracy of data fusion outputs. The vertical accuracy of the LiDAR point cloud is better than ±20 cm and its density is higher than 1 point per square meter. Some preprocessing steps are implemented to prepare LiDAR data for the accuracy assessment in different experiments. Details are explained in corresponding sections.
- *Building footprints*: The building footprints layer of the study area is provided by OpenStreetMap. The footprints layer is used in combination with heights derived from different sources for LOD1 3D reconstruction

4.1. Input DEM Generated by TanDEM-X and Cartosat-1 DEM Fusion

The first input data we used for LOD1 building model reconstruction, is a refined DEM resulting from a fusion of Cartosat-1 and TanDEM-X DEMs. As mentioned in Table 1, Cartosat-1 tiles are registered to highly accurate airborne orthophoto images to compensate horizontal misalignment. Before launching the TanDEM-X mission, Cartosat-1 tiles were vertically aligned with SRTM DEM as an almost global, open DEM. However, due to limited vertical accuracy of SRTM, TanDEM-X data can be substituted for vertical bias compensation of Cartosat-1 products. Thus, the alignment improves the vertical accuracy of the Cartosat-1 DEM. The evaluation illustrates that the absolute vertical accuracy of Cartosat-1 DEM increased more than 2 m. The evaluations were performed with respect to a LiDAR DSM created from the LiDAR point cloud by reducing and interpolating the 3D points into a 2.5D grid with a pixel spacing of 5 m. It should be noted that the TanDEM-X raw DEM is also converted into a 5 m pixel spacing DEM by interpolation. As we were able to show in [24], this fusion improves the final DEM quality; quantitative results for the test scene are repeated in Table 4.

Table 4. Accuracy (in meter) of Cartosat-1 and TanDEM-X DEM fusion in the urban study subset over Munich. The bold values indicate the best results which were obtained through the proposed DEM fusion pipeline.

DEM		Mean	RMSE	STD
Raw DEM	Cartosat-1	−0.68	5.27	5.23
	TanDEM-X	−**0.36**	6.43	6.42
Fused DEM	ANN-based	−0.55	**5.02**	**4.98**

4.2. Input DEM Generated by TanDEM-X Raw DEM Fusion

In the TanDEM-X mission, at least two primary DEMs are produced over all landmass tiles to reach the target relative accuracy [36]. This is realized by data fusion techniques such as weighted averaging. However, the weighted averaging performance is not optimal over urban areas. Therefore, in [27] we proposed to use efficient variational methods such as TV-L$_1$ and Huber models for fusing raw DEMs. We improved the height precision of the applied TanDEM-X raw DEM by employing another available tile (see Table 2). For this purpose, both TanDEM-X DEMs are converted to DEMs with pixel spacing of 6 m. The fusion performances using weighted averaging and variational models are shown in Figure 7. The quantitative results are collected in Table 5. Those evaluations are carried out with respect to a LiDAR DEM with 6 m pixel spacing achieved from the input LiDAR point cloud by interpolation.

Table 5. Height accuracy (in meters) of the TanDEM-X data before and after DEM fusion in the study area over Munich. The bold values indicate the best results which obtained through the TV-L_1-based fusion.

DEM		Mean	RMSE	STD
Fused DEM	WA	0.84	7.51	7.46
	TV-L_1	**0.77**	**6.11**	**6.06**
	Huber	0.78	6.14	6.09

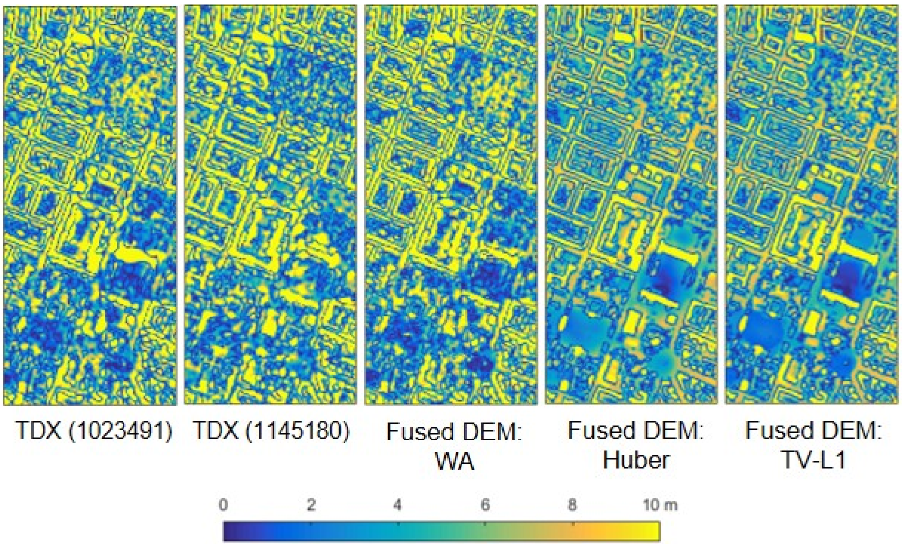


Figure 7. Absolute residual maps of the initial input raw DEMs and the fused DEMs obtained by different approaches for the study area over Munich.

As illustrated in Figure 7 and Table 5, the fusion can improve the quality of TanDEM-X raw DEMs. It becomes apparent that variational models, especially TV-L$_1$, outperform conventional weighted averaging model.

4.3. Input Point Cloud Generated by SAR-Optical Stereogrammetry

In [29], we have shown that by implementing a SAR-optical stereogrammetry framework for the TerraSAR-X and WorldView-2 image pairs, a sparse point cloud can be produced as a product of cooperative data fusion. A stereogrammetrically generated point cloud using MI as a similarity measure is shown in Figure 8.

To validate the accuracy of the resulting 3D point clouds, we employed the accurate airborne LiDAR point cloud described in Section 4. For accuracy calculation, after Least Square (LS) plane fitting on k (here: $k = 6$ points) nearest neighbors of each target point in the reference point cloud [37], the Euclidean distance between the target point to the fitted reference plane was measured along different directions. Table 6 summarizes accuracy assessments of the reconstructed point clouds using MI 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.

Table 6. 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	0.00	−0.04	0.27	1.57	1.69	3.09	1.57	1.69	3.10	2.75

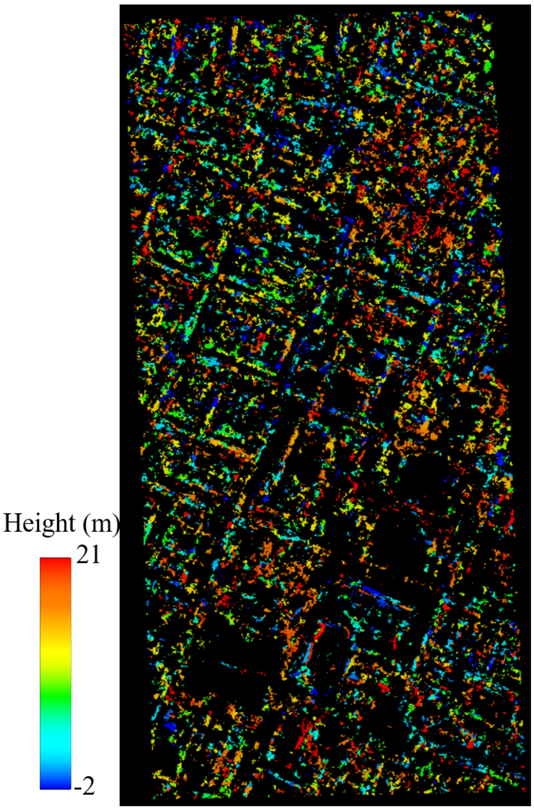


Figure 8. Achieved point cloud from stereogrammetric 3D reconstruction of TerraSAR-X/WorldView-2 over the Munich study subset.

5. Result of LOD1 Building Model Reconstruction

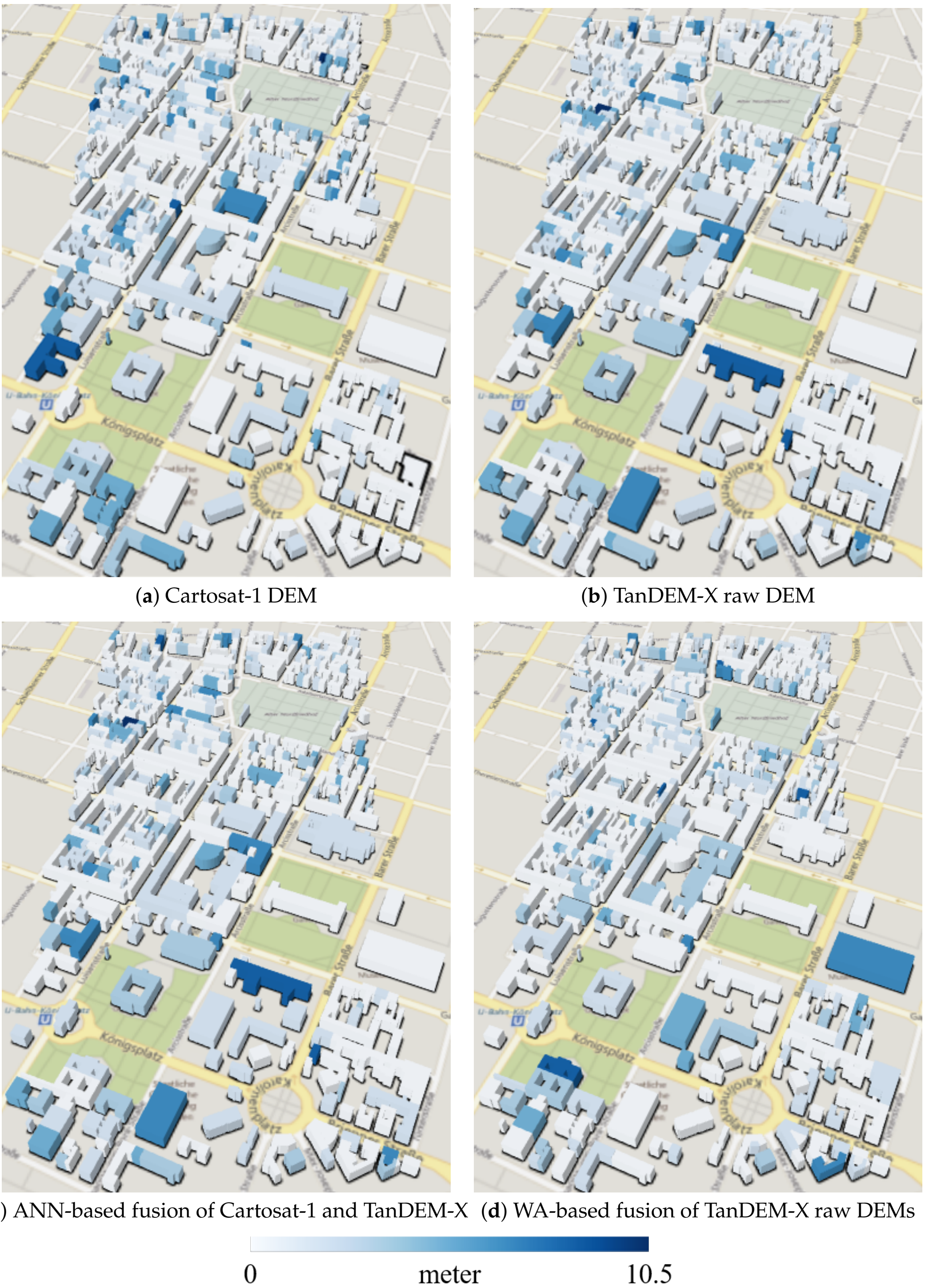
Figure 9 displays LOD1 3D reconstruction results for the study area consisting of prismatic building models generated by combining the height information derived from different sources discussed in the previous sections and building footprints provided by OpenStreetMap. As displayed in Figure 9, on average, all models are systematically biased in comparison to a model produced from high-resolution LiDAR data. However, this bias becomes minimum for a model using heights derived from SAR-optical stereogrammetry, as can be seen when comparing large buildings. However, for better evaluation, quantitative assessment should be performed. Therefore, the height accuracy of each LOD1 model was validated by comparing it with a model was created from the reference LiDAR DSM in a similar manner. For that purpose, we first interpolated the original LiDAR point cloud to a grid with a 1 m pixel spacing. Then, we used TV-L$_1$ denoising [27] to reduce potential noise effects. This TV-L$_1$ denoising mitigates biases in building height estimation induced by height outliers and inconsistencies such as those caused by crane-towers. As described in [27], TV-L$_1$ comprises two terms: a fidelity term and a penalty term. The effect of each term on the final output can be tuned by regularization parameters as weighting factors. Using a higher weight devoted to the penalty term will lead to better edge-preservation. Thus, we used the double weight for the penalty term to enhance urban structures. Then, the final height estimate within each building outline can be computed according to the process described in Section 3. The same process can be applied for the quality measurements of the 3D building reconstructions obtaining from other height information sources. The quantitative evaluations for the LOD1 reconstructions implemented based on scenario 1 (using original OSM) and 2 (using updated outlines) are presented in Tables 7 and 8, respectively.

Table 7. Quantitative evaluations (in meters) of the level-of-detail 1 (LOD1) reconstructions of the urban scene using heights derived from different sources along with original building outlines of OpenStreetMap (OSM).

Elevations		Median	RMSE	STD
input DEM	Cartosat-1	8.63	10.01	4.67
	TanDEM-X	9.68	10.16	4.28
Fused DEM	ANN-based: Cartosat-1 and TanDEM-X	9.56	9.97	4.28
	Weighted Averaging:TanDEM-X	7.91	9.5	4.81
	TV-L_1: TanDEM-X	8.94	8.95	3.82
	Huber: TanDEM-X	8.97	9	3.83
SAR-optical stereogrammetry	TerraSAR-X/WordlView-2	6.51	9.73	5.83

Table 8. Quantitative evaluations (in meters) of the LOD1 reconstructions of the urban scene using heights derived from different sources along with modified building outlines of OSM.

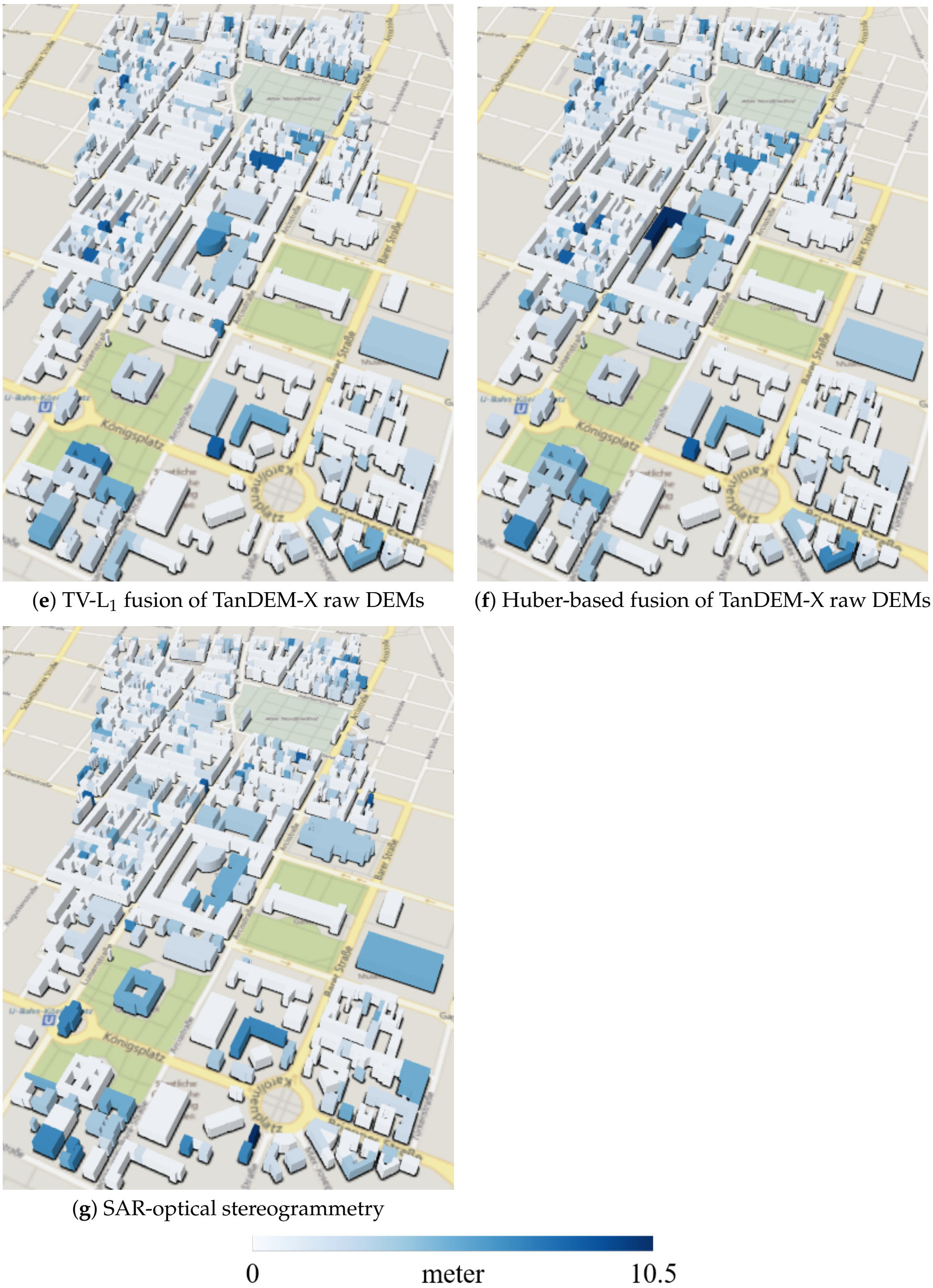
Elevations		Median	RMSE	STD
input DEM	Cartosat-1	−0.96	2.85	2.27
	TanDEM-X	−0.93	3.43	2.83
Fused DEM	ANN-based: Cartosat-1 and TanDEM-X	−0.92	3.09	2.48
	Weighted Averaging:TanDEM-X	−0.72	2.81	2.5
	TV-L_1: TanDEM-X	−0.68	2.86	2.56
	Huber: TanDEM-X	−0.67	2.96	2.64
SAR-optical stereogrammetry	TerraSAR-X/WorldView-2	−0.29	3.61	3.57



(**a**) Cartosat-1 DEM (**b**) TanDEM-X raw DEM

(**c**) ANN-based fusion of Cartosat-1 and TanDEM-X (**d**) WA-based fusion of TanDEM-X raw DEMs

Figure 9. *Cont.*



(**e**) TV-L_1 fusion of TanDEM-X raw DEMs

(**f**) Huber-based fusion of TanDEM-X raw DEMs

(**g**) SAR-optical stereogrammetry

Figure 9. Level-of-detail 1 (LOD1) reconstructions of the study urban scene using heights derived from different sources and building outlines obtained from building foot prints layer of OpenStreetMap (OSM). Colors indicate absolute height residuals.

6. Discussion

6.1. Multi-Sensor Fusion for Height Exploitation

In this research, we employed different sensor fusion techniques to use heights as a requirement for 3D building reconstruction. Two categories of techniques were used to improve the quality of TanDEM-X DEM as a global DEM. In the first method, using Cartosat-1 DEM could improve the quality of TanDEM-X. During DEM fusion, the issue of low absolute localization accuracy of Cartosat-1 DEM could be solved. It is also recommended to use TanDEM-X as an external DEM during the Cartosat-1 DEM generation to compensate bias existing in the sensor geometry. As a drawback, the Cartosat-1 data is not globally available such as TanDEM-X. Furthermore, due to different natures of TanDEM-X and Cartosat-1 DEMs, we implemented an ANN-based algorithm which utilizes both feature engineering and supervised training for weight map prediction. The weight maps are used for weighted averaging-based fusion to integrate TanDEM-X and Cartosat-1 DEMs. Nevertheless, the training samples do not necessarily exist in an arbitrary study area. The next possibility is to use other TanDEM-X covers acquired through the mission to guarantee target relative accuracy. For this, we implemented variational models to smooth noise appearing in DEMs while preserving the building outlines. The main advantage of variational techniques is that they do not need highly accurate training samples such as those derived from LiDAR data. In addition, it only employs TanDEM-X raw DEM tiles and does not require a higher quality DEM such as that derived from Cartosat-1 data. However, by comparing quantitative results represented in Tables 4 and 5 using different metrics, it is demonstrated that the first solution i.e., employing Cartosat-1 DEM and implementing ANN-based DEM fusion could ultimately generate a more accurate urban DEM.

Another opportunity for producing heights is to carry out stereogrametry for 3D reconstruction from archived SAR-optical image pairs such as TerraSAR-X and WorldView-2 images. The promising outputs demonstrated potential and possibility of 3D reconstruction from SAR-optical stereogrammetry. However, some development such as improving dense matching performance to produce a denser point cloud as well as noisy point and outlier removal are demanded.

6.2. LOD1 Building Reconstruction

After implementing data fusion techniques for height retrieval, we reconstructed building models using the derived heights and the building outlines provided by OSM. The achieved model is not a complete 3D city model since it provides building heights only. However, this model can be used for applications that require the building volume, which is not affected by the lack of information on the precise elevations of the building bottom/top. We investigated the reconstruction using original building outlines provided by OSM as well as using an updated building footprints layer. Regarding the median values in Table 7, using the original building outlines causes a bias affecting estimated final heights (RMSE values) while standard deviations are much smaller, thus confirming a systematic change in building heights. This bias can be significantly reduced by modifying building outlines in a preprocessing step (Table 8).

Using heights derived from outputs of multi-sensor DEM fusion can still lead to better reconstruction results in comparison to the primary TanDEM-X DEM. While the highest accuracy is obtained by Cartosat-1 data, it owes the accuracy to the bias compensation through the alignment to TanDEM-X. Without the alignment, the existing bias would be propagated to the final building heights.

Last but not least, it has to be mentioned that for generating a complete 3D city model, computing the height of the bottom and the top of a building along with the underlying terrain is required. Due to the limited the resolution of the height data utilized in this study, our focus did not lie on full 3D city model reconstruction but on simple prismatic building model reconstruction. For that purpose, we worked with the assumption of flat terrain at a constant height, which is valid in the selected study area. For a complete 3D city model, more accurate measurements of the terrain and the bottom of building elevations would be necessary.

7. Conclusions

In this research, we evaluated the potential of LOD1 3D reconstruction using data from remote-sensing-derived geodata and volunteered geographic information (VGI). For this purpose, we used heights derived from sources provided for global mapping such as those produced through the TanDEM-X mission. We implemented two DEM fusion experiments to improve the quality of TanDEM-X in urban areas. First is to fuse the TanDEM-X and Cartosat-1 DEMs using corresponding weight maps generated through a supervised ANN-based pipeline. In the second experiment, multiple TanDEM-X raw DEMs are fused by variational models. The results confirm the quality improvement of TanDEM-X after DEM fusion. In another experiment, heights were from an archived TerraSAR-X and WorldView-2 image pair through a stereogrammetry framework. The output was a sparse point cloud with a promising accuracy. Since building outlines as an essential requirement for 3D reconstruction cannot be accurately recognized in those height sources, we employed outlines provided by OSM. It was also shown that the primary outlines are not perfect and should be modified and updated for an accurate reconstruction. The final results demonstrate the possibility of prismatic building model generation (at LOD1 level) on a wide area from easily accessible, remote sensing-derived geodata.

Author Contributions: Conceptualization, Hossein Bagheri, Michael Schmitt, and Xiaoxiang Zhu; Methodology, Hossein Bagheri; Software, Hossein Bagheri; Data Curation and Investigation, Hossein Bagheri; Writing Hossein Bagheri; Review and editing, Michael Schmitt, and Xiaoxiang Zhu; Supervision, Michael Schmitt, and Xiaoxiang Zhu; Project Administration, Michael Schmitt; Funding Acquisition, Hossein Bagheri, Michael Schmitt; Resources, Xiaoxiang Zhu.

Funding: This work is jointly supported by the German Research Foundation (DFG) under grant SCHM 3322/1-1, the Helmholtz Association under the framework of the Young Investigators Group SiPEO (VH-NG-1018), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. ERC-2016-StG-714087, Acronym: So2Sat.

Acknowledgments: The authors would like to thank Donaubauer of the Chair of Geoinformatics of TUM for fruitful discussions about the CityGML standard and levels-of-detail for building models. In addition, they 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, the Bavarian Surveying Administration for the LiDAR reference data of Munich, Fritz and Baier of DLR for providing TanDEM-X raw DEMs, Reinartz and d'Angelo of DLR for providing the Cartosat-1 DEM.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Biljecki, F.; Stoter, J.; Ledoux, H.; Zlatanova, S.; Çöltekin, A. Applications of 3D City Models: State of the Art Review. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2842–2889. [CrossRef]
2. Kolbe, T.H.; Gröger, G.; Plümer, L. CityGML: Interoperable Access to 3D City Models. In *Geo-Information for Disaster Management*; van Oosterom, P., Zlatanova, S., Fendel, E.M., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 883–899.
3. Biljecki, F.; Ledoux, H.; Stoter, J. An improved LOD specification for 3D building models. *Comput. Environ. Urban Syst.* **2016**, *59*, 25–37.
4. Ledoux, H.; Meijers, M. Topologically consistent 3D city models obtained by extrusion. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 557–574. [CrossRef]
5. Biljecki, F.; Ledoux, H.; Stoter, J.; Vosselman, G. The variants of an LOD of a 3D building model and their influence on spatial analyses. *ISPRS J. Photogram. Remote Sens.* **2016**, *116*, 42–54. [CrossRef]
6. Kim, C.; Habib, A.; Chang, Y.C. Automatic generation of digital building models for complex structures from LiDAR data. *Int. Arch. Photogram. Remote Sens.* **2008**, *37*, 456–462.
7. Buyukdemircioglu, M.; Kocaman, S.; Isikdag, U. Semi-Automatic 3D City Model Generation from Large-Format Aerial Images. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 339. [CrossRef]
8. Gamba, P.; Houshmand, B.; Saccani, M. Detection and extraction of buildings from interferometric SAR data. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 611–617. [CrossRef]

9. Biljecki, F.; Ledoux, H.; Stoter, J. Generating 3D city models without elevation data. *Comput. Environ. Urban Syst.* **2017**, *64*, 1–18. [CrossRef]
10. Gröger, G.; Kolbe, T.H.; Nagel, C.; Häfele, K.H. OGC City Geography Markup Language (CityGML) Encoding Standard. Available online: https://www.opengeospatial.org/standards/citygml (accessed on 19 January 2019).
11. Stoter, J.; Vosselman, G.; Dahmen, C.; Oude Elberink, S.; Ledoux, H. CityGML Implementation Specifications for a Countrywide 3D Data Set. *Photogram. Eng. Remote Sens.* **2014**, *80*, 1069–1077. [CrossRef]
12. Arefi, H.; Engels, J.; Hahn, M.; Mayer, H. Levels of Detail in 3D Building Reconstruction from LiDAR Data. *ISPRS Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci.* **2008**, *XXXVII-B3b*, 485–490.
13. Kolbe, T.H.; Burger, B.; Cantzler, B. *CityGML Goes to Broadway*; Photogrammetric Week '15: Stuttgart, Germany, 2015; pp. 343–356.
14. Stoter, J.; Roensdorf, C.; Home, R.; Capstick, D.; Streilein, A.; Kellenberger, T.; Bayers, E.; Kane, P.; Dorsch, J.; Woźniak, P.; et al. 3D Modelling with National Coverage: Bridging the Gap Between Research and Practice. In *3D Geoinformation Science: The Selected Papers of the 3D GeoInfo 2014*; Breunig, M., Al-Doori, M., Butwilowski, E., Kuper, P.V., Benner, J., Haefele, K.H., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 207–225.
15. Rajpriya, N.; Vyas, A.; Sharma, S. Generation of 3D Model for Urban area using Ikonos and Cartosat-1 Satellite Imageries with RS and GIS Techniques. *Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 899–906. [CrossRef]
16. Marconcini, M.; Marmanis, D.; Esch, T.; Felbier, A. A novel method for building height estmation using TanDEM-X data. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 14–18 July 2014; pp. 4804–4807.
17. Nghiem, S.; Balk, D.; Small, C.; Deichmann, U.; Wannebo, A.; Blom, R.; Sutton, P.; Yetman, G.; Chen, R.; Rodriguez, E.; et al. Global Infrastructure: The Potential of SRTM Data to Break New Ground. Available online: https://www.researchgate.net/publication/228538455_Global_Infrastructure_The_Potential_of_SRTM_Data_to_Break_New_Ground (accessed on 19 January 2019).
18. Gamba, P.; Dell Acqua, F.; Houshmand, B. SRTM data characterization in urban areas. *Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci.* **2002**, *34*, 55–58.
19. Quartulli, M.; Datcu, M. Information fusion for scene understanding from interferometric SAR data in urban environments. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1976–1985. [CrossRef]
20. Misra, P.; Avtar, R.; Takeuchi, W. Comparison of Digital Building Height Models Extracted from AW3D, TanDEM-X, ASTER, and SRTM Digital Surface Models over Yangon City. *Remote Sens.* **2018**, *10*, 2008. [CrossRef]
21. Schmitt, M.; Zhu, X.X. Data Fusion and Remote Sensing: An ever-growing relationship. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 6–23. [CrossRef]
22. Srivastava, P.K.; Srinivasan, T.; Gupta, A.; Singh, S.; Nain, J.S.; Prakash, S.; Kartikeyan, B.; Krishna, B.G. Recent Advances in CARTOSAT-1 Data Processing. Available online: https://www.researchgate.net/publication/242118849_Recent_advances_in_Cartosat-1_data_processing (accessed on 19 January 2019).
23. Lehner, M.; Müller, R.; Reinartz, P.; Schroeder, M. Stereo evaluation of Cartosat-1 data for French and Catalonian test sites. In Proceedings of the ISPRS Hannover Workshop 2007: High Resolution Earth Imaging for Geospatial Information, Hannover, Germany, 2–5 June 2009.
24. Bagheri, H.; Schmitt, M.; Zhu, X.X. Uncertainty assessment and weight map generation for efficient fusion of TanDEM-X and Cartosat-1 DEMs. *ISPRS Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci.* **2017**, *XLII-1/W1*, 433–439. doi:10.5194/isprs-archives-XLII-1-W1-433-2017. [CrossRef]
25. Bagheri, H.; Schmitt, M.; Zhu, X.X. Fusion of TanDEM-X and Cartosat-1 elevation data supported by neural network-predicted weight maps. *ISPRS J. Photogram. Remote Sens.* **2018**, *144*, 285–297. [CrossRef]
26. Gruber, A.; Wessel, B.; Martone, M.; Roth, A. The TanDEM-X DEM Mosaicking: Fusion of Multiple Acquisitions Using InSAR Quality Parameters. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2016**, *9*, 1047–1057. [CrossRef]
27. Bagheri, H.; Schmitt, M.; Zhu, X.X. Fusion of Urban TanDEM-X Raw DEMs Using Variational Models. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2018**, *11*, 4761–4774. [CrossRef]
28. Wegner, J.D.; Ziehn, J.R.; Soergel, U. Combining High-Resolution Optical and InSAR Features for Height Estimation of Buildings With Flat Roofs. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5840–5854. [CrossRef]

29. Bagheri, H.; Schmitt, M.; d'Angelo, P.; Zhu, X.X. A Framework for SAR-Optical Stereogrammetry over Urban Areas. *ISPRS J. Photogram. Remote Sens.* **2018**, *146*, 389–408. [CrossRef] [PubMed]
30. Hirschmüller, H. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [CrossRef] [PubMed]
31. Hirschmuller, H.; Scharstein, D. Evaluation of Stereo Matching Costs on Images with Radiometric Differences. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 1582–1599. [CrossRef]
32. D2.8.III.2 INSPIRE Data Specification on Buildings—Technical Guidelines. Technical Report, European Commission Joint Research Centre. 2013. Available online: https://inspire.ec.europa.eu/id/document/tg/bu (accessed on 19 January 2019).
33. d'Angelo, P.; Lehner, M.; Krauss, T.; Hoja, D.; Reinartz, P. Towards Automated DEM Generation from High Resolution Stereo Satellite Images. *ISPRS Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci.* **2008**, *37*, 1137–1342.
34. The Federal Agency for Cartography and Geodesy of Germany (BKG). Digital Orthophotos. Available online: https://www.bkg.bund.de/SharedDocs/Downloads/BKG/DE/Downloads-DE-Flyer/AdV-DOP-DE (accessed on 17 September 2018).
35. Deledalle, C.; Denis, L.; Tupin, F. Iterative Weighted Maximum Likelihood Denoising With Probabilistic Patch-Based Weights. *IEEE Trans. Image Process.* **2009**, *18*, 2661–2672. [CrossRef] [PubMed]
36. Rizzoli, P.; Martone, M.; Gonzalez, C.; Wecklich, C.; Tridon, D.B.; Bräutigam, B.; Bachmann, M.; Schulze, D.; Fritz, T.; Huber, M.; et al. Generation and performance assessment of the global TanDEM-X digital elevation model. *ISPRS J. Photogram. Remote Sens.* **2017**, *132*, 119–139. [CrossRef]
37. Mitra, N.J.; Nguyen, A.; Guibas, L. Estimating surface normals in noisy point cloud data. *Int. J. Comput. Geom. Appl.* **2004**, *14*, 261–276. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Bibliography

[1] F. Biljecki, J. Stoter, H. Ledoux, S. Zlatanova, and A. Çöltekin. "Applications of 3D city models: state of the art review". In: *ISPRS International Journal of Geo-Information* 4.4 (2015), pp. 2842–2889 (cit. on p. 11).

[2] T. H. Kolbe, G. Gröger, and L. Plümer. "CityGML: interoperable access to 3D city models". In: *Geo-information for Disaster Management*. Ed. by P. van Oosterom, S. Zlatanova, and E. M. Fendel. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 883–899 (cit. on p. 11).

[3] Karlsruhe Institute of Technology (KIT), Institute for Applied Computer Science. *CityGML example FZK-Haus*. http://www.citygmlwiki.org/index.php?title=FZK_Haus. (Accessed 10.18). 2018 (cit. on p. 11).

[4] "The variants of an LOD of a 3D building model and their influence on spatial analyses". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 116 (2016), pp. 42 –54 (cit. on p. 11).

[5] C. Kim, A. Habib, and Y.-C. Chang. "Automatic generation of digital building models for complex structures from LiDAR data". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37.B4 (2008), pp. 456–462 (cit. on pp. 11, 13).

[6] M. Buyukdemircioglu, S. Kocaman, and U. Isikdag. "Semi-automatic 3D city model generation from large-format aerial images". In: *ISPRS International Journal of Geo-Information* 7.9 (2018) (cit. on pp. 11, 13).

[7] P. Gamba, B. Houshmand, and M. Saccani. "Detection and extraction of buildings from interferometric SAR data". In: *IEEE Transactions on Geoscience and Remote Sensing* 38.1 (2000), pp. 611–617 (cit. on pp. 11, 13).

[8] F. Biljecki, H. Ledoux, and J. Stoter. "Generating 3D city models without elevation data". In: *Computers, Environment and Urban Systems* (2017), pp. 1–18 (cit. on p. 11).

[9] H. Bagheri, M. Schmitt, and X. X. Zhu. "Uncertainty assessment and weight map generation for efficient fusion of TanDEM-X and Cartosat-1 DEMs". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42-1/W1 (2017), pp. 433–439 (cit. on pp. 12, 15, 37).

[10] G. Krieger, A. Moreira, H. Fiedler, I. Hajnsek, M. Werner, M. Younis, and M. Zink. "TanDEM-X: a satellite formation for high-resolution SAR interferometry". In: *IEEE Transactions on Geoscience and Remote Sensing* 45.11 (2007), pp. 3317–3341 (cit. on pp. 12, 23, 27).

[11] N. Rajpriya, A. Vyas, and S. Sharma. "Generation of 3D model for urban area using Ikonos and Cartosat-1 satellite imageries with RS and GIS techniques". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 40.8 (2014), pp. 899–906 (cit. on p. 14).

[12] M. Marconcini, D. Marmanis, T. Esch, and A. Felbier. "A novel method for building height estmation using TanDEM-X data". In: *2014 IEEE Geoscience and Remote Sensing Symposium*. 2014, pp. 4804–4807 (cit. on p. 14).

[13] S. Nghiem, D. Balk, C. Small, U. Deichmann, A. Wannebo, R. Blom, P. Sutton, G. Yetman, R. Chen, E. Rodriguez, et al. "Global infrastructure: the potential of SRTM data to break new ground". In: *White Paper Produced by CIESIN and NASA's Jet Propulsion Laboratory* (2001) (cit. on p. 14).

[14] P. Gamba, F. D. Acqua, and B. Houshmand. "SRTM data characterization in urban areas". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 34.3/B (2002), pp. 55–58 (cit. on p. 14).

[15] M. Quartulli and M. Datcu. "Information fusion for scene understanding from interferometric SAR data in urban environments". In: *IEEE Transactions on Geoscience and Remote Sensing* 41.9 (2003), pp. 1976–1985 (cit. on p. 14).

[16] P. Misra, R. Avtar, and W. Takeuchi. "Comparison of digital building height models extracted from AW3D, TanDEM-X, ASTER, and SRTM digital surface models over Yangon city". In: *Remote Sensing* 10.12 (2018) (cit. on p. 14).

[17] M. Schmitt and X. X. Zhu. "Data fusion and remote sensing: An ever-growing relationship". In: *IEEE Geoscience and Remote Sensing Magazine* 4.4 (2016), pp. 6–23 (cit. on pp. 14, 20, 22).

[18] M. Crosetto and B. Crippa. "Optical and radar data fusion for DEM generation". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 32.4 (1998), pp. 128–134 (cit. on p. 14).

[19] H. Schultz, E. M. Riseman, F. R. Stolle, and D.-M. Woo. "Error detection and DEM fusion using self-consistency". In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 2. 1999, pp. 1174–1181 (cit. on p. 14).

[20] P. Reinartz, R. Müller, D. Hoja, M. Lehner, and M. Schroeder. "Comparison and fusion of DEM derived from SPOT-5 HRS and SRTM data and estimation of forest heights". In: *Earsel Symposium, Porto, Portugal, 6.-11. June 2005*. Ed. by Earsel. 2005, on CD ROM (cit. on p. 14).

[21] A. Roth, W. Knöpfle, G. Strunz, M. Lehner, and P. Reinartz. "Towards a global elevation product: combination of multi-source digital elevation models". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 34.4 (2002), pp. 675–679 (cit. on p. 15).

[22] M. Costantini, F. Malvarosa, E. Minati, and E. Zappitelli. "A data fusion algorithm for DEM mosaicking: building a global DEM with SRTM-X and ERS data". In: *2006 IEEE International Symposium on Geoscience and Remote Sensing*. 2006, pp. 3861–3864 (cit. on p. 15).

[23] J. Hoffmann and D. Walter. “How complementary are SRTM-X and -C band digital elevation models?” In: *Photogrammetric Engineering & Remote Sensing* 72.3 (2006), pp. 261–268 (cit. on p. 15).

[24] C. Rossi, M. Eineder, T. Fritz, P. d’Angelo, and P. Reinartz. “Quality assessment of TanDEM-X raw DEMs oriented to a fusion with Cartosat-1 DEMs”. In: *33rd EARSeL Symposium*. 2013, pp. 1–9 (cit. on p. 15).

[25] R. Deo, C. Rossi, M. Eineder, T. Fritz, and Y. S. Rao. “Framework for fusion of ascending and descending pass TanDEM-X raw DEMs”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.7 (2015), pp. 3347–3355 (cit. on p. 15).

[26] A. Gruber, B. Wessel, M. Martone, and A. Roth. “The TanDEM-X DEM mosaicking: fusion of multiple acquisitions using InSAR quality parameters”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.3 (2016), pp. 1047–1057 (cit. on pp. 15, 23, 51).

[27] H. Bagheri, M. Schmitt, and X. X. Zhu. “Urban TanDEM-X Raw DEM Fusion Based on TV-L1 and Huber Models”. In: *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. 2018, pp. 7251–7254 (cit. on p. 15).

[28] H. Papasaika, E. Kokiopoulou, E. Baltsavias, K. Schindler, and D. Kressner. “Fusion of digital elevation models using sparse representations”. In: *Proceedings of the 2011 ISPRS Conference on Photogrammetric Image Analysis*. PIA’11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 171–184 (cit. on p. 15).

[29] H. Papasaika, D. Poli, and E. Baltsavias. “Fusion of digital elevation models from various data sources”. In: *2009 International Conference on Advanced Geographic Information Systems Web Services*. 2009, pp. 117–122 (cit. on pp. 15, 37).

[30] C. Zach, T. Pock, and H. Bischof. “A globally optimal algorithm for robust TV-L1 range image integration”. In: *2007 IEEE 11th International Conference on Computer Vision*. 2007, pp. 1–8 (cit. on pp. 15, 42).

[31] T. Pock, L. Zebedin, and H. Bischof. “TGV-fusion”. In: *Rainbow of Computer Science: Dedicated to Hermann Maurer on the Occasion of His 70th Birthday*. Ed. by C. S. Calude, G. Rozenberg, and A. Salomaa. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 245–258 (cit. on pp. 15, 42, 43).

[32] G. Kuschk, P. d’Angelo, D. Gaudrie, P. Reinartz, and D. Cremers. “Spatially regularized fusion of multi-resolution digital surface models”. In: *IEEE Transactions on Geoscience and Remote Sensing* 55.3 (2017), pp. 1477–1488 (cit. on pp. 16, 27).

[33] C. E. Fuss, A. A. Berg, and J. B. Lindsay. “DEM fusion using a modified k-means clustering algorithm”. In: *International Journal of Digital Earth* 9.12 (2016), pp. 1242–1255 (cit. on p. 16).

[34] P. Reinartz, R. Müller, P. Schwind, S. Suri, and R. Bamler. “Orthorectification of VHR optical satellite data exploiting the geometric accuracy of TerraSAR-X data”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 66.1 (2011), pp. 124–132 (cit. on p. 16).

[35] T. Toutin and L. Gray. “State-of-the-art of elevation extraction from satellite SAR data”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 55.1 (2000), pp. 13 –33 (cit. on p. 16).

[36] S. Méric, F. Fayard, and É. Pottier. “Radargrammetric SAR image processing”. In: *Geoscience and remote sensing*. InTech, 2009 (cit. on p. 16).

[37] K. Gutjahr, R. Perko, J. Raggam, and M. Schardt. “The epipolarity constraint in stereo-radargrammetric DEM generation”. In: *IEEE Transactions on Geoscience and Remote Sensing* 52 (2014), pp. 5014–5022 (cit. on pp. 16, 29).

[38] A. L. Bloom, E. J. Fielding, and X.-Y. Fu. “A demonstration of stereophotogrammetry with combined SIR-B and Landsat TM images”. In: *International Journal of Remote Sensing* 9.5 (1988), pp. 1023–1038 (cit. on p. 16).

[39] J. Raggam and A. Almer. “Mathematical aspects of multi-sensor stereo mapping”. In: *10th Annual International Symposium on Geoscience and Remote Sensing*. 1990, pp. 1963–1966 (cit. on p. 17).

[40] J. Raggam, A. Almer, and D. Strobl. “Multisensor mapping using SAR in conjunction with optical data”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 29 (1993), pp. 556–556 (cit. on p. 17).

[41] J. Raggam, A. Almer, and D. Strobl. “A combination of SAR and optical line scanner imagery for stereoscopic extraction of 3D data”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 49.4 (1994), pp. 11–21 (cit. on p. 17).

[42] S. Xing, Q. Xu, Y. Zhang, Y. He, and G. Jin. “Optical/SAR sensors stereo positioning”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37.B1 (2008), pp. 993–996 (cit. on p. 17).

[43] J. D. Wegner, J. R. Ziehn, and U. Soergel. “Combining high-resolution optical and InSAR features for height estimation of buildings with flat roofs”. In: *IEEE Transactions on Geoscience and Remote Sensing* 52.9 (2014), pp. 5840–5854 (cit. on p. 17).

[44] M. Schmitt and X. X. Zhu. “On the challenges in stereogrammetric fusion of SAR and optical imagery for urban areas”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 41-B7 (2016), pp. 719–722 (cit. on p. 17).

[45] C. Qiu, M. Schmitt, and X. X. Zhu. “Towards automatic SAR-optical stereogrammetry over urban areas using very high resolution imagery”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 138 (2018), pp. 218–231 (cit. on p. 17).

[46] *D2.8.III.2 inspire data specification on buildings – technical guidelines*. Tech. rep. (Accessed 01.19). `https://inspire.ec.europa.eu/id/document/tg/bu`: European Commission Joint Research Centre, 2013 (cit. on pp. 19, 20).

[47] F. Biljecki, H. Ledoux, and J. Stoter. “Height references of CityGML LOD1 buildings and their influence on applications”. In: *Proceedings. 9th ISPRS 3D GeoInfo Conference 2014*, Dubai, UAE, Nov. 2014 (cit. on p. 19).

[48] A. Pope, W. G. Rees, A. J. Fox, and A. Fleming. "Open access data in polar and cryospheric remote sensing". In: *Remote Sensing* 6.7 (2014), pp. 6183–6220 (cit. on p. 21).

[49] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi. "Multisensor data fusion: a review of the state-of-the-art". In: *Information Fusion* 14.1 (2013), pp. 28–44 (cit. on p. 20).

[50] B. Rabus, M. Eineder, A. Roth, and R. Bamler. "The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 57.4 (2003), pp. 241–262 (cit. on p. 22).

[51] E. Rodriguez, C. S. Morris, and J. E. Belz. "A global assessment of the SRTM performance". In: *Photogrammetric Engineering & Remote Sensing* 72.3 (2006), pp. 249–260 (cit. on p. 22).

[52] T. Tachikawa, M. Kaku, A. Iwasaki, D. B. Gesch, M. J. Oimoen, Z. Zhang, J. J. Danielson, T. Krieger, B. Curtis, J. Haase, et al. *ASTER global digital elevation model version 2-summary of validation results.* Tech. rep. NASA, 2011 (cit. on p. 22).

[53] J. Takaku, T. Tadono, and K. Tsutsui. "Generation of high resolution global DSM from ALOS PRISM". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 40.4 (2014), pp. 243–248 (cit. on p. 22).

[54] T. Tadono, H. Ishida, F. Oda, S. Naito, K. Minakawa, and H. Iwamoto. "Precise global DEM generation by ALOS PRISM". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2.4 (2014), pp. 71–76 (cit. on p. 22).

[55] J. Takaku, T. Tadono, K. Tsutsui, and M. Ichikawa. "Validation of "AW3D" global DSM generated from ALOS PRISM". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 3 (2016), pp. 25–31 (cit. on p. 23).

[56] B. Heady, G. Kroenung, and C. Rodarmel. "High-resolution elevation data (HRE) specification overview". In: *ASPRS/MAPPS 2009 Conference, San Antonio, Texas.* Nov. 2009, on CD ROM (cit. on p. 23).

[57] P. Rizzoli, M. Martone, C. Gonzalez, C. Wecklich, D. B. Tridon, B. Bräutigam, M. Bachmann, D. Schulze, T. Fritz, M. Huber, B. Wessel, G. Krieger, M. Zink, and A. Moreira. "Generation and performance assessment of the global TanDEM-X digital elevation model". In: *ISPRS Journal of Photogrammetry and Remote Sens.* 132 (2017), pp. 119–139 (cit. on pp. 23, 24).

[58] T. Fritz, C. Rossi, N. Yague-Martinez, F. Rodriguez-Gonzalez, M. Lachaise, and H. Breit. "Interferometric processing of TanDEM-X data". In: *2011 IEEE International Geoscience and Remote Sensing Symposium.* 2011, pp. 2428–2431 (cit. on p. 23).

[59] T. Fritz, H. Breit, C. Rossi, U. Balss, M. Lachaise, and S. Duque. "Interferometric processing and products of the TanDEM-X mission". In: *2012 IEEE International Geosci. and Remote Sens. Symposium.* 2012, pp. 1904–1907 (cit. on p. 23).

[60] J. H. González, J. M. W. Antony, M. Bachmann, G. Krieger, M. Zink, D. Schrank, and M. Schwerdt. "Bistatic system and baseline calibration in TanDEM-X to ensure the global digital elevation model quality". In: *ISPRS Journal of Photogrammetry and Remote Sens.* 73 (2012), pp. 3–11 (cit. on p. 23).

[61] A. Gruber, B. Wessel, M. Huber, and A. Roth. "Operational TanDEM-X DEM calibration and first validation results". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 73 (2012), pp. 39–49 (cit. on p. 23).

[62] B. Wessel, U. Marschalk, A. Gruber, M. Huber, T. Hahmann, A. Roth, and M. Habermeyer. "Design of the DEM mosaicking and calibration processor for TanDEM-X". In: *7th European Conference on Synthetic Aperture RADAR.* 2008, pp. 1–4 (cit. on p. 23).

[63] M. Zink, M. Bachmann, B. Brautigam, T. Fritz, I. Hajnsek, A. Moreira, B. Wessel, and G. Krieger. "TanDEM-X: the new global DEM takes shape". In: *IEEE Geoscience and Remote Sensing Magazine* 2.2 (2014), pp. 8–23 (cit. on p. 23).

[64] C. Rossi and S. Gernhardt. "Urban DEM generation, analysis and enhancements using TanDEM-X". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 85 (2013), pp. 120–131 (cit. on p. 23).

[65] P. K. Srivastava, T. Srinivasan, A. Gupta, S. Singh, J. S. Nain, S. Prakash, B Kartikeyan, and B. G. Krishna. "Recent advances in Cartosat-1 data processing". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36-1/W51 (2007), on CD–ROM (cit. on p. 24).

[66] N. Ahmed, A. Mahtab, R. Agrawal, P. Jayaprasad, S. K. Pathan, Ajai, D. K. Singh, and A. K. Singh. "Extraction and validation of Cartosat-1 DEM". In: *Journal of the Indian Society of Remote Sensing* 35.2 (2007), pp. 121–127 (cit. on p. 24).

[67] A. Uttenthaler, F. Barner, T. Hass, J. Makiola, P. d'Angelo, P. Reinartz, S. Carl, and K. Steiner. "Euro-maps 3D- a transnational, high-resolution digital surface model for Europe". In: *ESA Living Planet Symposium.* Vol. 722. 2013, p. 271 (cit. on p. 24).

[68] X. Hu. "Methods for quality control of large Cartosat-1 stereo blocks". Diploma. Universität Stuttgart, 2011 (cit. on pp. 24–26).

[69] M. Lehner, R. Müller, P. Reinartz, and M. Schroeder. "Stereo evaluation of Cartosat-1 data for French and Catalonian test sites". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36-1/W51 (2007), on CD–ROM (cit. on p. 24).

[70] T.-A. Teo. "Bias compensation in a rigorous sensor model and rational function model for high-resolution satellite images". In: *Photogrammetric Engineering & Remote Sensing* 77.12 (2011), pp. 1211–1220 (cit. on p. 24).

[71] T. Kim and J. Jeong. "DEM matching for bias compensation of rigorous pushbroom sensor models". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 66.5 (2011), pp. 692–699 (cit. on p. 24).

[72] P. d'Angelo, M. Lehner, T. Krauss, D. Hoja, and P. Reinartz. "Towards automated DEM generation from high resolution stereo satellite images". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37.B4 (2008), pp. 1137–1342 (cit. on p. 25).

[73] M. Ravanbakhsh and C. S. Fraser. "A comparative study of DEM registration approaches". In: *Journal of Spatial Science* 58.1 (2013), pp. 79–89 (cit. on p. 25).

[74] A. Gruen and D. Akca. "Least squares 3D Surface and curve matching". In: *ISPRS Journal of Photogrammetry and Remote Sens.* 59.3 (2005), pp. 151–174 (cit. on p. 25).

[75] W. Birgit. *TanDEM-X ground segment–DEM products specification document.* `https://elib.dlr.de/108014/1/TD-GS-PS-0021_DEM-Product-Specification_v3.1.pdf`. (Accessed 03.19). 2016 (cit. on p. 26).

[76] J. Höhle and M. Höhle. "Accuracy assessment of digital elevation models by means of robust statistical methods". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 64.4 (2009), pp. 398–406 (cit. on p. 26).

[77] M. Rumpler, A. Wendel, and H. Bischof. "Probabilistic range image integration for DSM and true-orthophoto generation". In: *Image Analysis*. Ed. by J.-K. Kämäräinen and M. Koskela. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 533–544 (cit. on p. 27).

[78] M. F. Morgan, N. El-Sheimy, and A. Saalfeld. "Epipolar resampling of linear array scanner scenes". PhD dissertation. University of Calgary, Department of Geomatics Engineering, 2004 (cit. on p. 28).

[79] W. Cho, T. Schenk, and M. Madani. "Resampling digital imagery to epipolar geometry". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 29 (1993), pp. 404–404 (cit. on p. 28).

[80] R. I. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Second. Cambridge University Press, 2004 (cit. on p. 28).

[81] R. Gupta and R. I. Hartley. "Linear pushbroom cameras". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.9 (1997), pp. 963–975 (cit. on p. 29).

[82] M. J. V. Zoej and G. Petrie. "Mathematical modelling and accuracy testing of SPOT level 1B stereopairs". In: *The Photogrammetric Record* 16.91 (1998), pp. 67–82 (cit. on p. 29).

[83] C. Lee, H. J. Theiss, J. S. Bethel, and E. M. Mikhail. "Rigorous mathematical modeling of airborne pushbroom imaging systems". In: *Photogrammetric Engineering and Remote Sensing* 66.4 (2000), pp. 385–392 (cit. on p. 29).

[84] T. Kim. "A study on the epipolarity of linear pushbroom images". In: *Photogrammetric Engineering and Remote Sensing* 66.8 (2000), pp. 961–966 (cit. on p. 29).

[85] A. B. Orun. "A modified bundle adjustment software for SPOT imagery and photography: tradeoff". In: *Photogrammetric Engineering and Remote Sensing* 60.12 (1994), pp. 1431–1437 (cit. on p. 29).

[86] M. Morgan, K. Kim, S. Jeong, and A. Habib. “Epipolar geometry of linear array scanners moving with constant velocity and constant attitude”. In: *ISPRS, Istanbul, Turkey* 1024 (2004), on CD ROM (cit. on p. 29).

[87] J. C. Curlander. “Location of spaceborne SAR imagery”. In: *IEEE Transactions on Geoscience and Remote Sensing* GE-20.3 (1982), pp. 359–364 (cit. on pp. 29, 32).

[88] D. Li and Y. Zhang. “A rigorous SAR epipolar geometry modeling and application to 3D target reconstruction”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6.5 (2013), pp. 2316–2323 (cit. on p. 29).

[89] J. Grodecki and G. Dial. “Block adjustment of high-resolution satellite images described by rational polynomials”. In: *Photogrammetric Engineering & Remote Sensing* 69.1 (2003), pp. 59–68 (cit. on pp. 29, 32, 60).

[90] X. Tong, S. Liu, and Q. Weng. “Bias-corrected rational polynomial coefficients for high accuracy geo-positioning of QuickBird stereo imagery”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.2 (2010), pp. 218–226 (cit. on p. 29).

[91] C. S. Fraser and H. B. Hanley. “Bias-compensated RPCs for sensor orientation of high-resolution satellite imagery”. In: *Photogrammetric Engineering & Remote Sensing* 71.8 (2005), pp. 909–915 (cit. on p. 29).

[92] M. Z. Brown, D. Burschka, and G. D. Hager. “Advances in computational stereo”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25.8 (2003), pp. 993–1008 (cit. on p. 30).

[93] D. Scharstein, R. Szeliski, and R. Zabih. “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms”. In: *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*. 2001, pp. 131–140 (cit. on p. 30).

[94] H. Hirschmüller. “Stereo processing by semiglobal matching and mutual information”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.2 (2008), pp. 328–341 (cit. on p. 30).

[95] V. Kratky. “Rigorous stereophotogrammetric treatment of SPOT images”. In: *CNES, SPOT 1 Image Utilization, Assessment, Results* (1988), pp. 1281–1288 (cit. on p. 31).

[96] J. C. Curlander and R. N. McDonough. *Synthetic aperture radar*. Vol. 396. John Wiley & Sons New York, NY, USA, 1991 (cit. on p. 32).

[97] J. Oh, W. H. Lee, C. K. Toth, D. A. Grejner-Brzezinska, and C. Lee. “A piecewise approach to epipolar resampling of pushbroom satellite images based on RPC”. In: *Photogrammetric Engineering & Remote Sensing* 76.12 (2010), pp. 1353–1363 (cit. on p. 32).

[98] C. Fraser, G. Dial, and J. Grodecki. “Sensor orientation via RPCs”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 60.3 (2006), pp. 182–194 (cit. on p. 32).

[99] R. Li, F. Zhou, X. Niu, and K. Di. “Integration of Ikonos and QuickBird imagery for geopositioning accuracy analysis”. In: *Photogrammetric Engineering and Remote Sensing* 73.9 (2007), p. 1067 (cit. on p. 32).

[100] C. V. Tao and Y. Hu. "3D reconstruction methods based on the rational function model". In: *Photogrammetric Engineering & Remote Sensing* 68.7 (2002), pp. 705–714 (cit. on p. 32).

[101] C. V. Tao, Y. Hu, and W. Jiang. "Photogrammetric exploitation of Ikonos imagery for mapping applications". In: *International Journal of Remote Sensing* 25.14 (2004), pp. 2833–2853 (cit. on p. 32).

[102] T. Toutin. "Comparison of 3D physical and empirical models for generating DSMs from stereo HR images". In: *Photogrammetric Engineering & Remote Sensing* 72.5 (2006), pp. 597–604 (cit. on p. 32).

[103] C. Tao and Y. Hu. "Use of the rational function model for image rectification". In: *Canadian Journal of Remote Sensing* 27.6 (2001), pp. 593–602 (cit. on p. 32).

[104] J. Grodecki, G. Dial, and J. Lutes. "Mathematical model for 3D feature extraction from multiple satellite images described by RPCs". In: *ASPRS Annual Conference Proceedings, Denver, Colorado*. 2004, on CD ROM (cit. on p. 32).

[105] C. V. Tao and Y. Hu. "A comprehensive study of the rational function model for photogrammetric processing". In: *Photogrammetric engineering and remote sensing* 67.12 (2001), pp. 1347–1358 (cit. on p. 32).

[106] L. Zhang, X. He, T. Balz, X. Wei, and M. Liao. "Rational function modeling for spaceborne SAR datasets". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 66.1 (2011), pp. 133–145 (cit. on p. 33).

[107] C. Rossi, T. Fritz, H. Breit, and M. Eineder. "First urban TanDEM-X raw DEMs analysis". In: *2011 Joint Urban Remote Sensing Event*. 2011, pp. 65–68 (cit. on p. 36).

[108] H. Bagheri, M. Schmitt, and X. X. Zhu. "Fusion of TanDEM-X and Cartosat-1 elevation data supported by neural network-predicted weight maps". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 144 (2018), pp. 285–297 (cit. on p. 36).

[109] H. Bagheri, M. Schmitt, and X. X. Zhu. "Fusion of urban TanDEM-X raw DEMs using variational models". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.12 (2018), pp. 4761–4774 (cit. on pp. 36, 66).

[110] T. Toutin. "Impact of terrain slope and aspect on radargrammetric DEM accuracy". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 57.3 (2002), pp. 228–240 (cit. on p. 37).

[111] P. Reinartz, P. d'Angelo, T. Krauß, D. Poli, K. Jacobsen, and G. Buyuksalih. "Benchmarking and quality analysis of DEM generated from high and very high resolution optical stereo satellite data". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (2010) (cit. on p. 37).

[112] V. Olaya. "Chapter 6: basic land-surface parameters". In: *Geomorphometry Concepts, Software, Applications*. Vol. 33. Elsevier, 2009, pp. 141–169 (cit. on p. 37).

[113] L. Birgé and Y. Rozenholc. "How many bins should be put in a regular histogram". In: *ESAIM: Probability and Statistics* 10 (2006), pp. 24–45 (cit. on p. 38).

[114] L. I. Rudin, S. Osher, and E. Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D: Nonlinear Phenomena* 60.1-4 (1992), pp. 259–268 (cit. on pp. 42, 43).

[115] M. Nikolova. "A variational approach to remove outliers and impulse noise". In: *Journal of Mathematical Imaging and Vision* 20.1 (2004), pp. 99–120 (cit. on p. 42).

[116] M. Bertero, C. De Mol, and G. A. Viano. "The Stability of Inverse Problems". In: *Inverse Scattering Problems in Optics*. Ed. by H. P. Baltes. Berlin, Heidelberg: Springer Berlin Heidelberg, 1980, pp. 161–214 (cit. on p. 43).

[117] T. F. Chan and S. Esedoglu. "Aspects of total variation regularized L1 function approximation". In: *SIAM Journal on Applied Mathematics* 65.5 (2005), pp. 1817–1837 (cit. on p. 43).

[118] T. Chan, S. Esedoglu, F. Park, and A. Yip. *Total variation image restoration: overview and recent developments*. Ed. by N. Paragios, Y. Chen, and O. Faugeras. Boston, MA: Springer US, 2006, pp. 17–31 (cit. on p. 43).

[119] P. J. Huber. "Robust statistics". In: *International Encyclopedia of Statistical Science*. Ed. by M. Lovric. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1248–1251 (cit. on p. 43).

[120] R. Perko and C. Zach. "Globally optimal robust DSM fusion". In: *European Journal of Remote Sens.* 49.1 (2016), pp. 489–511 (cit. on p. 43).

[121] A. Chambolle and T. Pock. "A first-order primal-dual algorithm for convex problems with applications to imaging". In: *Journal of Mathematical Imaging and Vision* 40.1 (2011), pp. 120–145 (cit. on p. 44).

[122] A. Thiele, E. Cadario, K. Schulz, U. Thoennessen, and U. Soergel. "Modeling and analyzing InSAR phase profiles at building locations". In: *2007 IEEE International Geoscience and Remote Sensing Symposium*. 2007, pp. 5053–5056 (cit. on p. 45).

[123] H. Bagheri, M. Schmitt, P. d'Angelo, and X. X. Zhu. "A framework for SAR-Optical stereogrammetry over urban areas". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 146 (2018), pp. 389–408 (cit. on pp. 55, 60).

[124] USGS. *Shuttle Radar Topography Mission (SRTM) Void Filled*. `https://lta.cr.usgs.gov/SRTMVF`. (accessed 09.17). USGS, 2000 (cit. on p. 61).

[125] S. Suri and P. Reinartz. "Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas". In: *IEEE Transactions on Geoscience and Remote Sensing* 48.2 (2010), pp. 939–949 (cit. on p. 61).

[126] G. Egnal and R. P. Wildes. "Detecting binocular half-occlusions: empirical comparisons of five approaches". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.8 (2002), pp. 1127–1133 (cit. on p. 61).

[127] N. J. Mitra, A. Nguyen, and L. Guibas. "Estimating surface normals in noisy point cloud data". In: *International Journal of Computational Geometry & Applications* 14 (2004), pp. 261–276 (cit. on p. 64).

[128] H. Bagheri, M. Schmitt, and X. Zhu. “Fusion of Multi-Sensor-Derived Heights and OSM-Derived Building Footprints for Urban 3D Reconstruction”. In: *ISPRS International Journal of Geo-Information* 8.4 (2019) (cit. on p. 65).

List of figures

List of tables

Acknowledgments

First and foremost, I would like to thank Dr. Schmitt who supervised and supported me throughout my Ph.D. life. I would also like to thank my supervisor, Prof. Zhu, for her feedbacks, comments, and advice on my investigations.

I also want to thank Dr. d'Angelo, for supervising me during the implementation of SAR-optical stereogrammetry research as well as for providing the Cartosat-1 DEM of Munich so I could apply it in the DEM fusion investigation. I would also like to thank Dr. Fritz and Dr. Baier for providing TanDEM-X data. In addition, I need to thank Prof. Reinartz for providing an opportunity and facilities for doing research at DLR.

I also had the pleasure of working alongside great people in the same situation as me. All of them contributed to or eased, in some way or other, the work on my thesis.

I also need to thank Prof. Sadeghian, who supervised my master's thesis and put me on the track of pursuing a Ph.D. in the first place. I also appreciate the assistance of Dr. Arefi for starting my Ph.D. course.

I need to thank Andre Richter, Michael Vonbun, Christian Herber, Stefan Wallentowitz, and Andreas Oeldemann for their Latex template, which made writing this thesis so much easier.

Finally, I am deeply grateful to my wife, Ms. Khalili, for encouraging and supporting me in a self-less manner throughout my professional career from master's to Ph.D. I am also indebted to my little son who provided sweet occasions in my life and lifted my spirits.