*Article*

# Mapping with Pléiades—End-to-End Workflow

**Roland Perko [1,2,]*[ID], Hannes Raggam [1] and Peter M. Roth [3][ID]**

[1]  JOANNEUM RESEARCH Forschungsgesellschaft mbH, DIGITAL, Remote Sensing and Geoinformation, 8010 Graz, Austria
[2]  Institute of Geodesy, Graz University of Technology, 8010 Graz, Austria
[3]  Institute of Computer Graphics and Vision, Graz University of Technology, 8010 Graz, Austria
[*]  Correspondence: roland.perko@joanneum.at

check for updates

**Abstract:**  In this work, we introduce an end-to-end workflow for very high-resolution satellite-based mapping, building the basis for important 3D mapping products: (1) digital surface model, (2) digital terrain model, (3) normalized digital surface model and (4) ortho-rectified image mosaic. In particular, we describe all underlying principles for satellite-based 3D mapping and propose methods that extract these products from multi-view stereo satellite imagery. Our workflow is demonstrated for the Pléiades satellite constellation, however, the applied building blocks are more general and thus also applicable for different setups. Besides introducing the overall end-to-end workflow, we need also to tackle single building blocks: optimization of sensor models represented by rational polynomials, epipolar rectification, image matching, spatial point intersection, data fusion, digital terrain model derivation, ortho rectification and ortho mosaicing. For each of these steps, extensions to the state-of-the-art are proposed and discussed in detail. In addition, a novel approach for terrain model generation is introduced. The second aim of the study is a detailed assessment of the resulting output products. Thus, a variety of data sets showing different acquisition scenarios are gathered, allover comprising 24 Pléiades images. First, the accuracies of the 2D and 3D geo-location are analyzed. Second, surface and terrain models are evaluated, including a critical look on the underlying error metrics and discussing the differences of single stereo, tri-stereo and multi-view data sets. Overall, 3D accuracies in the range of 0.2 to 0.3 m in planimetry and 0.2 to 0.4 m in height are achieved w.r.t. ground control points. Retrieved surface models show normalized median absolute deviations around 0.9 m in comparison to reference LiDAR data. Multi-view stereo outperforms single stereo in terms of accuracy and completeness of the resulting surface models.

**Keywords:** Pléiades satellites; multi-view stereo; RPC sensor model; accuracy analysis; 3D mapping; 3D reconstruction; digital surface model; digital terrain model; ortho rectification

## 1. Introduction

For many applications in remote sensing, highly accurate and up-to-date mapping information gathered from very high resolution (VHR) satellite stereo images is needed. Mapping, in general, can be described as the process of generating 3D maps of a region of interest on the Earth's surface consisting of 3D coordinates linked with the spectral information that was observed by the underlying sensor. Exemplary applications in the domain of remote sensing are, for instance, city modeling [1–3], forest assessment and biomass estimation [4–6], change detection [7,8], land cover and land use classification [9,10], carbon reporting [11], farm land monitoring [12], glacier observation [13,14], disaster damage mapping [15,16] and mapping in general [17,18]. To allow semantic analysis those applications need mapping products in form of digital surface models (DSM), digital terrain models (DTM), their difference, that is, normalized digital surface models (nDSM) and the according

multi-spectral ortho-rectified images. All this information can be extracted from stereo or multi-view satellite imagery. Therefore, this work focuses on the generation of those 3D mapping products on the example of images gathered from the Pléiades satellite constellation. Scientifically, this work is based upon our previous publications [19–21] which are extended, refined and completed. The central contributions of this work can be summarized as follows:

- An end-to-end workflow for mapping with satellite data is defined, especially tailored to the Pléiades sensors. Each component of this workflow is reviewed in detail.
- The specific implementation is described, where many insights are presented which often focus on efficient implementation yielding speed-up.
- Many extensions of the classical stereo processing chain are explained, including epipolar rectification, stereo matching details, fusion methods and also assessment details.
- A novel, robust and fast DSM to DTM conversion algorithm is presented that is very beneficial for various applications.
- For the first time in remote sensing, a multiple view geometry data set is processed comprising six Pléiades images that form 15 stereo pairs in along and across track direction. Note that other multiple view combinations were already processed in, for example, References [22–25].
- A comprehensive assessment is performed based on 24 Pléiades images gathered with a variety of acquired scenarios and distributed over the globe. In particular, assessments of the accuracy of 2D and 3D geo-location, such as DSMs and DTMs, are performed. Additional, results on stereo, tri-stereo and multi-view data sets are analyzed in detail.
- Directions of future research are pointed out, such as finding optimized matching cost functions or automatic ground control points (GCP) transfer from synthetic aperture radar (SAR) data to optical images which will then complement the whole workflow.

The special interest in using Pléiades data to generate those needed mapping products comes from the fact that this sensor can acquire tri-stereo panchromatic images in one single over flight (single pass along track stereo capacity) in high spatial resolution. In addition, several works show that it is possible to derive highly accurate DSMs from such stereo or tri-stereo images [19,21,26,27]. Note, that all presented and proposed methodologies in this work are not limited to this specific Pléiades satellites but can be directly applied to any satellite imagery with rational polynomial coefficient based sensor models. For instance, for Ikonos, Spot-7, Spot-8, WorldView-3, WorldView-4 and also the upcoming Pléiades-Neo constellation. The presented methods are all implemented within the commercial software package *Remote Sensing Software Graz (RSG)* (http://www.remotesensing.at/en/remote-sensing-software.html) and thus are available for the remote sensing community.

*1.1. The Pléiades Satellite System*

The European Pléiades satellite system is a dual system comprising the two identical satellites Pléiades-1A and Pléiades-1B. They have been launched in December 2011 and in December 2012, respectively, both providing VHR image data (where VHR is defined for resolutions or ground sampling distances (GSD) below 1 m [28]). The satellites operate in the same orbit with an offset of 180 degrees to offer a daily revisit capacity. They also share the same orbit as the satellites Spot-6 and Spot-7 but are positioned 90 degrees phase shifted. They are supplied with remarkable agility, as the pointing angles can be triggered in a range of ±47 degrees (standard mode ±30 degrees). The sensors are capable of acquiring a panchromatic band (470–830 nm) with 0.7 m GSD at nadir and four multi-spectral bands (blue: 430–550 nm; green: 500–620 nm; red: 590–710 nm; near-infrared: 740–940 nm) with 2.8 m GSD. Images of the Pléiades sensor get delivered as a bundle of a panchromatic band upsampled to 0.5 m GSD and multi-spectral bands at 2.0 m GSD in GeoTIFF or JPEG format. The multi-spectral information can be pansharpenend [29,30] or, alternatively, a multi-spectral pansharpend product can directly be ordered. The swath width of Pléiades image data is 20 km on ground, that corresponds to 40,000 pixels in across track direction. Interestingly, the Pléiades

open star cluster was the best place to steer the eponymous satellite in the calibration phase [31]. For additional information on the Pléiades sensors we refer to References [27,32–35].

The satellite agility is of importance with respect to 3D mapping from Pléiades image data. Similar to the present VHR missions of Ikonos, Spot or WorldView, stereo data can be acquired during one overflight (single pass) through an appropriate forward and backward arrangement of the sensor. A significant innovation and advantage of Pléiades, however, is provided through the capability to acquire even three images for an area, taken from the same orbit at along track forward-, nadir- and backward-view of the sensor and through the possibility of an across track swipe. Such image triplets are also denoted as tri-stereo data sets. An assessment of the benefit of such arrangements was made, for example, in Reference [36], with respect to image triplets composed from multi-sensor and multi-temporal acquisitions, respectively. In addition to that, the Pléiades sensors are also able to steer in across track direction such that they can collect images over the same scene on ground from different orbits yielding also across track stereo pairs. This enables the collection of multiple view geometry data sets (term taken from Reference [37]) as well known from computer vision and photogrammetry. The sensor model is delivered as a rational polynomial coefficient (RPC) model. The geo-location accuracy of this model is reported to be in the order of 8.5 m CE90 [32]. According to Reference [38] CE90 is defined as the radial error which 90% of all errors in a circular distribution will not exceed, that is, the 90th percentile of circular error distribution. Therefore, for accurate mapping, the geo-location accuracy has to be refined, that is, optimized using GCPs. Additionally, the launch of the Pléiades-Neo satellite constellation is planned between 2020 and 2022 [39]. It consists of four satellites with a panchromatic band at 0.3 m GSD, six spectral bands (deep blue, blue, green, red, red edge, near-infrared) at 1.2 m GSD, a swath width of 14 km, an envisaged geo-location accuracy of 5 m CE90 and a mission lifetime of more than 10 years.

### 1.2. End-to-End 3D Mapping

This work proposes a complete end-to-end workflow necessary for 3D mapping, where the main goal is to automatically and robustly derive the mapping products from the input data. The input data consists of (1) multi-view satellite images, that can consist of one or multiple stereo pairs, one or multiple triplets or arbitrary combinations like one along track triplet and a single image from another orbit and (2) related metadata, which are the sensor model for each image, reference GCPs and their 2D image coordinates in each given image. Generally, the image coordinates are manually measured within an image-based interactive measurement tool. This information is then employed in an automatic workflow that is comprised of four core tasks sketched schematically in Figure 1. Those tasks are sensor model adjustment, DSM generation, DTM generation and ortho-rectification. The final output of the proposed process are a DSM, a DTM, an nDSM and an ortho-rectified image mosaic, that is, the envisaged 3D mapping products. The flow graph for those tasks is depicted in Figure 2 with the following logic and objectives:

1.  As the initial sensor models are too inaccurate for precise mapping (cf. Reference [21]), those models have to be optimized in the first task. As depicted in the flow graph in Figure 2, the sensor models of each image are optimized separately in the parameter adjustment by employing the initial sensor model together with the GCPs and the image coordinates. The adjusted sensor models are later used in DSM generation and ortho-rectification. Within this task the geo-location accuracy is assessed in detail, that is, assessment of a-priori accuracy as well as assessment of the need and feasibility for optimization, typically by means of GCPs and least squares adjustment. For several test sites and acquisition scenarios, an investigation regarding 2D and 3D accuracy assessment is given. Optimization necessities and potentials are demonstrated in this context.

2.  The second task takes all images and their adjusted sensor models as input and derives one final DSM. First, an intermediate DSM is generated from each possible stereo pair. As shown in Figure 2, we propose the following production chain: Epipolar rectification based on the adjusted sensor models followed by stereo matching that yields a disparity map. Spatial point intersection

then results in 3D coordinates (i.e., a 3D point cloud) that is resampled into an intermediate DSM. Second, those DSMs are fused and post-processed yielding the final DSM. Within this task, the characteristics of the underlying sensor system is exploited, like availability of image triplets in case of Pléiadesand including high-level processing techniques for the given input. The applicability of high-level processing options, like the semi-global matching (SGM) technique, is investigated, enhanced and resulting achievements are illustrated and validated.

3.   In the third task, one DTM and one nDSM are generated solely based on the DSM produced in the previous task. The main steps are ground point filtering, detecting points in the given DSM which are located on bare earth, hole filling to interpolate non-ground information and post-processing mainly for outlier removal. In this way, we can (a) filter man-made structures and vegetation and (b) retrieve vegetation or building heights, finally yielding the DTM and nDSM models. While this is a well-established procedure for airborne light detection and ranging (LiDAR) data, where first and last pulse data can be utilized, much higher complexity is given when terrain needs to be purely extracted from DSMs derived from satellite stereo data. In this respect, algorithmic developments have been made to adapt and improve yet existing approaches to retrieve reliable terrain from such input.

4.   The fourth task takes all images, their adjusted sensor models and the final DSM from task two as input and derives one final multi-spectral ortho-rectified image mosaic. First, each image is ortho-rectified based on direct rectification yielding intermediate ortho images. Second, those ortho-images are fused to one final ortho-mosaic within the mosaicing step.

The proposed workflow is designed such that multi-node processing can easily be implemented. Especially, all processes that run over each image or stereo pair (i.e., the green boxes in Figure 2) can be executed in parallel on multiple nodes. Also note, that Sections 2, 4 and 5 follow the structure of the proposed workflow given in Figures 1 and 2.

*1.3. Outline*

This paper is organized as follows: Related work is reviewed in detail in Section 2. Then, the test sites and test data are described in Section 3. Next, Section 4 reports on the methodologies proposed in this work, followed by the results given in Section 5. All insights gained are discussed in Section 6. Finally, conclusions are drawn in Section 7.
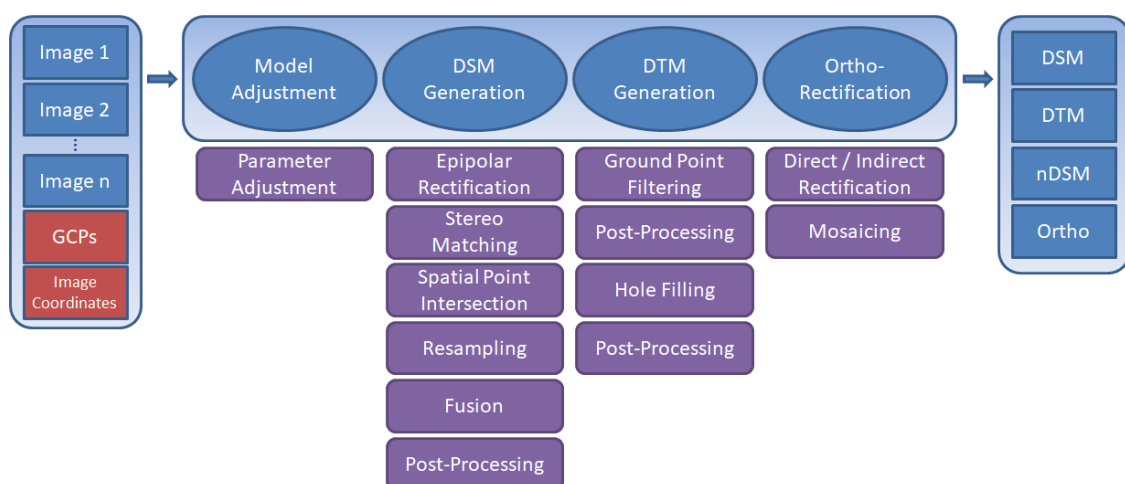


**Figure 1.** Proposed end-to-end 3D mapping workflow for multi-view stereo Pléiades images. The four core tasks are subdivided into according sub tasks.
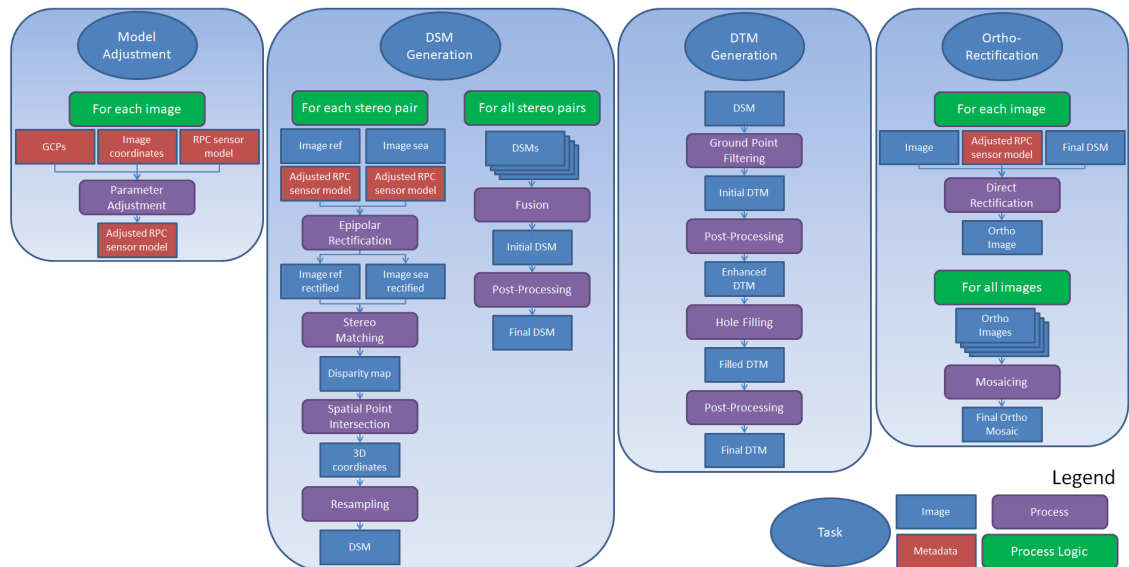
**Figure 2.** Proposed end-to-end 3D mapping flow graph for multi-view stereo Pléiades images. The four core tasks introduced in Figure 1 are depicted in detail.

## 2. Related Work

The proposed end-to-end workflow for 3D mapping with VHR Pléiades satellite stereo imagery touches various scientific independent topics that are reviewed in this section. Due to the goal of end-to-end mapping, those topics are intertwined like pieces of a puzzle, together bringing the complete processing chain into being. The presented topics are multidisciplinary and touch the fields of remote sensing, photogrammetry, computer vision and applied mathematics. This section follows the structure as given in Figure 1, thus touching the four core tasks of model adjustment, DSM generation, DTM generation and ortho-rectification, including all required sub-tasks.

### 2.1. Sensor Modeling and Parameter Optimization

For Pléiades images, a set of rational polynomial coefficients (RPC) is given, which define the rational polynomial sensor model and thus substitute a physical sensor model. This generic solution dates back to Reference [40] and for some sensors replaced the physical model. Many sensor models of optical VHR satellites are delivered with RPCs, such as Cartosat-2, CBERS-2, Deomis-2, EROS B, GaoFen-2, GeoEye-1, Ikonos, IRS-P5, Jilin-1, Kompsat-2, Kompsat-3, Kompsat-3A, QuickBird, Pléiades, RapidEye, SkySat, Spot-6, Spot-7, Spot-8, SuperView-1, TeLEOS-1, TripleSat, WorldView-3 or WorldView 4. Also some SAR sensor models, like for instance ERS-1, are supplied with RPCs. The main advantage of using the RPC model is that it presents a general solution for sensor modeling such that novel sensors are automatically supported by a given implementation. The disadvantage is that the coefficients are not directly related to physical properties like, for instance, sensor position or viewing direction. Therefore, in the adjustment procedure (cf. Reference [41]) it is hard to interpret which effect caused the change of coefficients. Mathematically, a non-linear equation system has to be solved to adjust the sensor parameters. This can be accomplished with numerical methods, such as Newton's method, Gauss-Newton algorithm, gradient descent or Levenberg-Marquardt algorithm. Depending on the number of released parameters and starting values the adjustment delivers robust solutions or gets stuck in a local minimum.

There is also an ongoing discussion on which parameters to be optimized. For instance, only a shift is considered in References [42,43] while linear coefficients are also considered in References [21,40,44,45].

The need to manually measure GCPs and their related image coordinates represent the main bottleneck of mapping. Therefore, automatic methods for GCP definition and transfer are of utter

importance. One idea is to employ SAR images, like from the TerraSAR-X sensor. This sensor provides extremely high geo-location accuracy [46] and thus perspectives to retrieve highly accurate 3D point information via radargrammetric processing [47–49]. Then such GCPs can be transferred into the optical Pléiades image via multi-modal image matching as presented, for instance, in References [50–52].

## 2.2. Digital Surface Model Generation

Accurate 3D reconstruction from stereo or multiple images is a well-known and well-researched topic in computer vision, photogrammetry and remote sensing. As depicted in Figure 1 the main steps are epipolar rectification, stereo matching, spatial point intersection, DSM resampling, DSM fusion and the related post-processing.

### 2.2.1. Epipolar Rectification

Epipolar rectification for push-broom images can be seen as a solved problem [53,54]. However, in contrast to perspective geometry, where analytic solutions exist for epipolar rectification (cf. References [55,56]), such rectification can only be approximated for push-broom images. The epipolar curves can be estimated by using the modified collinearity equations [57] and are hyperbola-like non-linear curves [58]. Some authors ignore the specific mathematical formulation of the push-broom collinear equations and treat the images as being perspective. These solutions work on a local level, that is, rectifying small parts of the images separately as shown in References [59,60]. In reality, it is sufficient if the location differences orthogonal to the epipolar direction (e.g., the column direction of the epipolar images) are below half a pixel [21]. Recently, it was also shown that an epipolar rectification with small residual errors can be achieved for high-resolution SAR satellite images [54]. Alternatively, epipolar rectification might be skipped if the subsequent stereo matching algorithm searches along the (non-linear) epipolar curves such as in Reference [61]. The drawbacks of this approach are that the image matching algorithm has to be general and has to consider the geometric constraints of the underlying sensor model. Additionally, for each location in the search space along the epipolar curve the local search window has to be resampled during matching.

### 2.2.2. Stereo Matching

Within the standard processing chain, the most challenging part is the correspondence problem, where in the dense case for each pixel coordinate in the first image the corresponding pixel coordinate in the second image which holds the same object has to be determined. The process of finding those parallaxes, shift vectors or disparities is called image matching. In principle, the matching problem determines the column and line shifts for each pixel and thus is based on a 2D search space. However, using the epipolar rectification for each pixel in the first image the corresponding pixel in the second image is located on a horizontal line. In this case, the search space in image matching becomes one-dimensional, thus enabling faster and more accurate matching algorithms. A very good benchmark of 1D matching methods can be found in Reference [62]. Having epipolar rectified images a standard procedure in photogrammetry and remote sensing is to apply the semi-global matching (SGM) introduced by Reference [61]. This matching scheme is able to achieve very good results, comparable to methods based on global optimization, for example, graph cut [63] or total variation [64]. Moreover, it is computationally very efficient. Currently many commercial packages for processing satellite data use epipolar rectification and SGM for DSM generation, for example, Erdas Imagine (http://www.hexagongeospatial.com/products/remote-sensing/erdas-imagine/overview), PCI Geomatica (http://www.pcigeomatics.com), Trimble Inpho with Match-T or MATCH-3DX (http://www.trimble.com/imaging/inpho.aspx), s2p (http://dev.ipol.im/~carlo/s2p/), SURE (http://nframes.com/) (only works with perspective images [65]), MicMac (http://micmac.ensg.eu) [66], RPC Stereo Processor (http://u.osu.edu/qin.324/rsp/) [67] and also the Remote Sensing Software Graz (http://www.remotesensing.at/en/remote-sensing-software.html). Recent research reports on

different variants of SGM, such as Reference [65] and extensions like those in References [68–71]. As underlying cost function most algorithms apply the Hamming distance of the Census transform [72] due to its robustness [73]. Alternatively, the so-called modified Census Transform [74,75] is used. Also a weighting can be incorporated based on the radiometric or the geometric distance (similarity and proximity weighting) [75,76]. Optimizing the stereo matching technique is always an important research topic. In particular, a combined cost-function of Census transform and absolute difference measure is envisaged but also the replacement of Census transform with generalized Census transform [77,78] is an issue. The most recent step is to train the cost-function via deep learning of a convolutional neural network [79–82].

### 2.2.3. Spatial Point Intersection

Using the sensor models and the disparities 3D points are extracted using spatial forward point intersection. For each matching result, there are four equations (two per image) to determine the 3D ground coordinates. The according non-linear equation system can be solved by linearization of the sensor model followed by a least-squares adjustment to find the coordinate increments. The spatial forward intersection leads to a 3D point cloud non-equally distributed in space [21].

### 2.2.4. DSM Resampling

There exist different paradigms, like storing the 3D information as point clouds [83,84] or as a DSM with given spacing. The first option is state-of-the-art in aerial photogrammetry while the second one is mostly used in satellite-based remote sensing. Additionally, in the second case, all information from a multi-stereo data set can simultaneously be used in the spatial point intersection to achieve a higher over-determination in the least-squares approach [36]. Another method is to extract DSMs from each stereo pair and then fuse those DSMs to a final DSM [85–87]. This method has the advantage that multiple stereo pairs can be processed independently in parallel (e.g., on multiple PCs), while only the fusion process has to access all data. There are several methods for mapping the 3D point cloud onto a DSM, for example, interpolation of a regular grid of height values and filling the remaining gaps by an appropriate interpolation mechanism [21] or by LiDAR-based approaches like that in Reference [88].

### 2.2.5. DSM Fusion

As mentioned above multiple DSMs stemming from individual stereo pairs (e.g., from a triple acquisition or neighboring stereo acquisitions) have to be fused into one final DSM. One option is to apply local approaches like mean, median or probability model-based filtering [85]. The other option is to state a global optimization problem to be solved numerically via total variation [87] or via gradient descent [86].

### 2.2.6. Post-Processing

Appropriate post-processing is always of interest, also at different stages of the presented workflow. For instance, occluded regions may be filled with an interpolation technique that filters towards the ground level. This avoids blurring of 3D breaklines while still gathering a high spatial reconstruction coverage. A post-processing option can also be seen in the so-called peak filter that segments images into radiometrically connected regions [61,89]. Small or isolated regions can then be discarded.

### 2.2.7. Additional Notes

In general, there exist workflows that can extract DSMs from Pléiades stereo or tri-stereo data. However, it is difficult to evaluate various approaches and to find the best solution for different applications. One good example is the Pléiades workshop held in Paris in November 2014 named *DEM computation from satellite images: Existing tools and developments* where different

solutions were benchmarked. The main challenge remains in finding the optimal parameters for image matching (like a custom-tailored cost function, appropriate occlusion detection and handling principle), while being computationally efficient (this includes memory optimizations, parallel processing on a multi-core central processing unit (CPU) and exhaustive usage of a graphical processing unit (GPU)). From an implementation point of view it would be beneficial to have an epipolar rectification method that can be applied for push-broom optical and SAR images.

### 2.3. Digital Terrain Model Generation

In the literature, DTM generation approaches are described for different kind of input data. Most of them are designated for airborne LiDAR data, some for DSMs extracted from high-resolution airborne optical images and spaceborne optical imagery. In this work, we aim for a DTM extraction approach custom-tailored to DSMs that are generated by a state-of-the-art photogrammetric workflow from VHR satellite stereo or tri-stereo images, as presented in Section 2.2. Thus, a novel algorithm will be proposed within this paper which is especially suitable for this case.

### 2.3.1. Airborne LiDAR Data

There are many publications on DTM generation from LiDAR data, where the first and last pulse data exist. Good review papers are [90,91]. In regions of vegetation, the first pulse measurement corresponds to the top height of the vegetation while the last pulse is reflected from bare-earth, thus defining the terrain height. Therefore, shrubs, trees and whole forests can be easily filtered based upon LiDAR full-pulse data. Man-made objects like buildings, however, are present in both LiDAR measurements and still have to be filtered by appropriate algorithms (or even manually) to get a correct DTM. Since LiDAR provides very accurate height measurements, 3D breaklines are in general very well preserved in LiDAR DSMs. When moving along a 1D profile from street level to a building a sharp steep height jump is observed, which can be used as a distinct feature for the DTM generation process (cf. Reference [92]). One representative algorithm is described in Reference [93], which will serve as the baseline for the algorithmic evolution presented in this paper. A method on hierarchical robust linear prediction employing LiDAR data is introduced in Reference [94].

### 2.3.2. Airborne Optical Data

For very high resolution airborne cameras the work [95] reports on a method developed for UltraCam images [96,97]. They perform a classification of the ortho images into 15 classes, both using texture and height information. Then the classification is employed within the DTM generation process. The main difference to spaceborne data is that the underlying DSMs are of significantly higher resolution and of higher accuracy such that image classification works very robust—which would not be the case for spaceborne data. A variational approach which uses a DSM as single input is presented in Reference [98]. The basic concept is to extract a very smooth surface using a strong regularization weight within a variational formulation. This over-smoothed surface is used to determine potential points the on ground (called detection mask) which are then used to interpolate a DTM by employing the same variational formulation as before. However, problems occur on large buildings that are detected as belonging to the ground. Thus, the authors proposed to add a segmentation based on maximally stable extremal regions and a manual interaction step.

### 2.3.3. Spaceborne Optical Data

For satellite-based stereo DSMs, three DTM generation approaches based on morphology, geodesic dilation and steep edge detection on simulated synthetic urban scenes are evaluated in Reference [99]. A standard method based on morphological opening was extended in Reference [100]. There, percentile filters rather than morphological grayvalue erosion and dilation are employed to be more robust against DSM outliers. Another approach is based on determining the normalized volume above ground over several scanlines [101]. An automated stratified object-based approach that also uses

the spectral information of an ortho image is presented in Reference [102]. An approach based on iterative geodesic reconstruction tested on Cartosat-1 stereo images is presented in Reference [103]. DTM generation in forested regions based on Cartosat-1 stereo image is reported in Reference [104]. There, the panchromatic image is classified into semantic classes, like buildings, low forest, high forest or ground and different filtering is applied to DSM regions according to the class.

*2.4. Ortho-Rectification and Mosaicing*

The process of ortho-rectification is well described in literature. The main two approaches are direct and indirect rectification. The first projects every input pixel into the resulting map product by an image-to-map transformation. The second starts from the map and projects every point back into the input image by a map-to-image transformation. The process of mosaicing, that is, stitching multiple individual ortho images to one product is understood as well. Design issues are the selection of pixels for one map cell and the feathering option. For both processes we refer to the standard literature [105–109]. A remaining challenge can be seen in an efficient integration into the complete workflow, which will be discussed in his work.

## 3. Test Sites and Data Sets

To allow sophisticated statements on Pléiades-based mapping, several test sites have been selected. They represent different acquisition dispositions in terms of viewing angles, stereo, triplet and multi-stereo constellations, cover different types of terrain and originate from both Pléiades-1A and 1B sensors. The test sites are located in Austria, Italy, Slovenia, Singapore and China. For the core investigations, three test sites were used, for which both representative Pléiades acquisitions and appropriate reference data sets are available. The data from Singapore must not be published, thus no screenshots of this test site are presented in this work. Overall, 24 Pléiades images are used in this study, with acquisition dates and main acquisition parameters as listed in Table 1. Images are also assigned with an ID such that referring to a specific image gets simpler. When talking about a stereo model it is addressed as two IDs, like model 1–2.

For assessment of the 2D and 3D geo-location accuracy reference GCPs are needed. In the optimal case, these points are measured with in-situ highly accurate devices, such as differential GPS. This was done for the test site Singapore. Alternatively, existing orthophotos with higher spatial resolution can be used to manually define the GCPs where the height information is taken from the LiDAR reference. This was done for the test site Ljubljana. In other cases, airborne LiDAR information is used to derive GCPs. There, distinctive locations, especially at 3D breaklines, are manually measured (for instance, a corner of a roof top or a hydrant). Obviously, as the LiDAR DSMs are available with 1 m GSD those manually measured objects will never get the accuracy as the in-situ or orthophoto measurements. This has to be kept in mind when comparing the accuracy values, as the accuracy of the reference information is the lower bound for the resulting geo-location accuracy.

The multitude of test sites were chosen to demonstrate that the proposed end-to-end workflow allows precise 3D mapping over the whole globe and is able to operate with different reference data and multi-stereo dispositions.

Since the incidence angles of Pléiades images are given for across track (roll, $\omega$), for along track (pitch, $\phi$) and for the combined overall angle, we derive equations to calculate the stereo intersection angle $\delta$ (also called the convergence angle [42]) between two Pléiades scenes and accordingly an approximation of their base-to-height (B/H) ratio. Thus, we first define the two vectors $r_1$ and $r_2$ and second they are used to compute $\delta$ and B/H as

$$\delta = \arccos\left(\frac{r_1 \cdot r_2}{|r_1| \cdot |r_2|}\right) \quad \text{and} \quad \frac{B}{H} = 2 \cdot \tan\frac{\delta}{2} \quad \text{with} \quad r_1 = \begin{bmatrix} \tan\phi_1 \\ \tan\omega_1 \\ 1 \end{bmatrix} \quad \text{and} \quad r_2 = \begin{bmatrix} \tan\phi_2 \\ \tan\omega_2 \\ 1 \end{bmatrix}. \quad (1)$$

All convergence angles within this work are determined using Equation (1), especially the ones given in Tables 1 and 2.

**Table 1.** Acquisition parameters for the Pléiades satellite images over the study areas (with $\delta$ being the convergence angle based on Equation (1)).

| Study Area | Acq. Date and Platform | ID | Time | Incidence Angle [°] | | | Model | $\delta$ [°] |
|---|---|---|---|---|---|---|---|---|
| | | | | Along | Across | Overall | | |
| Innsbruck Stereo | 2013-07-13 1A | 1 | 10:16:41.4 | −11.72 | 8.96 | 14.60 | 1–2 | 26.9 |
| | | 2 | 10:17:25.8 | 14.48 | 2.35 | 14.66 | | |
| Innsbruck Triplet East | 2014-06-09 1A | 1 | 10:21:06.6 | −18.74 | 4.11 | 19.13 | 1–2 | 10.1 |
| | | 2 | 10:21:23.8 | −8.99 | 1.42 | 9.09 | 2–3 | 23.3 |
| | | 3 | 10:22:02.0 | 13.61 | −4.63 | 14.32 | 1–3 | 33.4 |
| Innsbruck Triplet Center | 2014-06-08 1B | 1 | 10:29:07.4 | −17.82 | −8.53 | 19.53 | 1–2 | 24.9 |
| | | 2 | 10:29:49.3 | 6.54 | −15.46 | 16.67 | 2–3 | 4.6 |
| | | 3 | 10:29:57.1 | 11.15 | −16.77 | 19.80 | 1–3 | 29.5 |
| Innsbruck Triplet West | 2014-06-09 1A | 1 | 10:21:15.6 | −14.30 | 3.82 | 14.76 | 1–2 | 10.4 |
| | | 2 | 10:21:32.9 | −4.21 | 1.13 | 4.36 | 2–3 | 23.0 |
| | | 3 | 10:22:11.1 | 18.11 | −4.89 | 18.67 | 1–3 | 33.4 |
| Trento Triplet | 2012-09-11 1A | 1 | 10:14:23.5 | −16.86 | 14.68 | 21.83 | 1–2 | 5.5 |
| | | 2 | 10:14:33.3 | −11.25 | 13.34 | 17.20 | 2–3 | 27.3 |
| | | 3 | 10:15:18.8 | 15.74 | 7.00 | 17.09 | 1–3 | 32.9 |
| Ljubljana Stereo 1 | 2013-07-27 1A | 1 | 10:10:08.2 | −12.12 | −1.98 | 14.9 | 1–2 | 25.2 |
| | | 2 | 10:10:49.7 | 12.36 | −8.59 | 12.3 | | |
| Ljubljana Stereo 2 | 2013-07-28 1B | 3 | 10:03:42.1 | −7.39 | 9.36 | 15.3 | 3–4 | 22.7 |
| | | 4 | 10:04:19.6 | 14.78 | 3.95 | 11.8 | | |
| Ljubljana Stereo 3 | 2013-07-29 1A | 5 | 9:55:03.4 | −10.20 | 22.31 | 24.1 | 5–6 | 26.2 |
| | | 6 | 9:55:48.9 | 16.91 | 16.90 | 23.3 | | |
| Singapore Stereo | 2014-10-01 1B | 1 | 3:33:16.8 | −0.31 | 8.64 | 8.64 | 1–2 | 15.2 |
| | | 2 | 3:33:41.9 | 14.65 | 5.44 | 15.55 | | |
| Tian Shui Stereo | 2014-05-07 1A | 1 | 3:53:38.3 | −10.54 | 0.75 | 10.56 | 1–2 | 22.8 |
| | | 2 | 3:54:15.4 | 11.68 | −4.54 | 12.48 | | |

### 3.1. Test Site Innsbruck (Austria)

This test site covers urban, rural and mountainous terrain, with ellipsoidal heights ranging from 560 m to 2750 m and spans over an area of 1200 km$^2$. In July 2013, only one stereo pair was acquired from Pléiades-1A after the launch of Pléiades-1B and thus after the recalibration of the Pléiades-1A satellite. One year after, within two days in June 2014 three image triplets were achieved, two of them acquired by Pléiades-1A and one by Pléiades-1B , which show different geographic areas of this test site, that is, eastern, center and western part. The central part of the test site is shown in Figure 3. The triplets do not have an overlap on the ground, such that images from different triplets do not form across track stereo pairs. For Innsbruck, an airborne LiDAR DSM is available as reference data set, collected in 2006 at 2.0 points/m$^2$. The temporal gap between Pléiades and LiDAR data has to be taken into account in the evaluation since changes occurred due to urban build-up and vegetation grow or clear-cut. The LiDAR DSM is available as raster data in UTM 32 North map projection and WGS 84 datum at 1 m GSD with ellipsoidal heights and could be used further to measure GCPs. 3D discontinuities in the LiDAR data can be used to define and manually measure ground coordinates. A fairly high number of GCPs, homogeneously distributed in planimetry and height, is measured. Targets like road intersections, corners of houses, water bodies or field boundaries, serve as control point candidates. Thus, 30 points were acquired for Innsbruck.
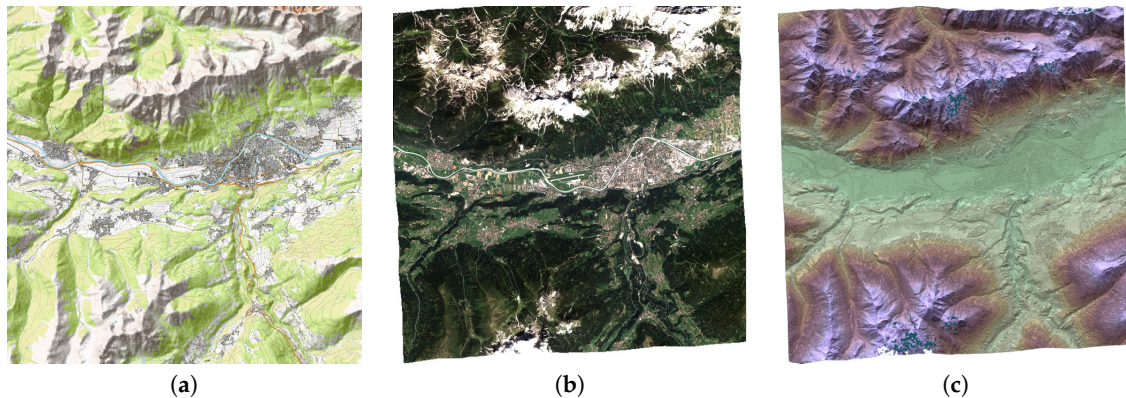
**Figure 3.** The test site Innsbruck (Austria): (**a**) topographic map (opentopomap.org (CC-BY-SA)), (**b**) RGB ortho-image and (**c**) relief shaded DSM. (**b**,**c**) were produced employing the proposed and implemented mapping workflow.

### 3.2. Test Site Trento (Italy)

This test site covers rural as well as mountainous terrain, the ellipsoidal heights ranging from 175 m to 1550 m and spans over an area of 220 km$^2$. A Pléiades-1A triplet was acquired in September 2012, which however is far from optimal for 3D reconstruction, since the first two images have a very small intersection angle of only 5.5°, while the second stereo pair has a huge intersection angle of 27.3°. The test site is shown in Figure 4. Also for Trento an airborne LiDAR DSM is available as reference data set, collected in 2007 at 1.3 points/m$^2$, available as raster data in UTM 32 North map projection and WGS 84 datum at 1 m GSD with ellipsoidal heights. Again there is a huge temporal gap between Pléiades and LiDAR acquisition to be taken into consideration. 21 GCPs were measured from LiDAR and serve as reference points.
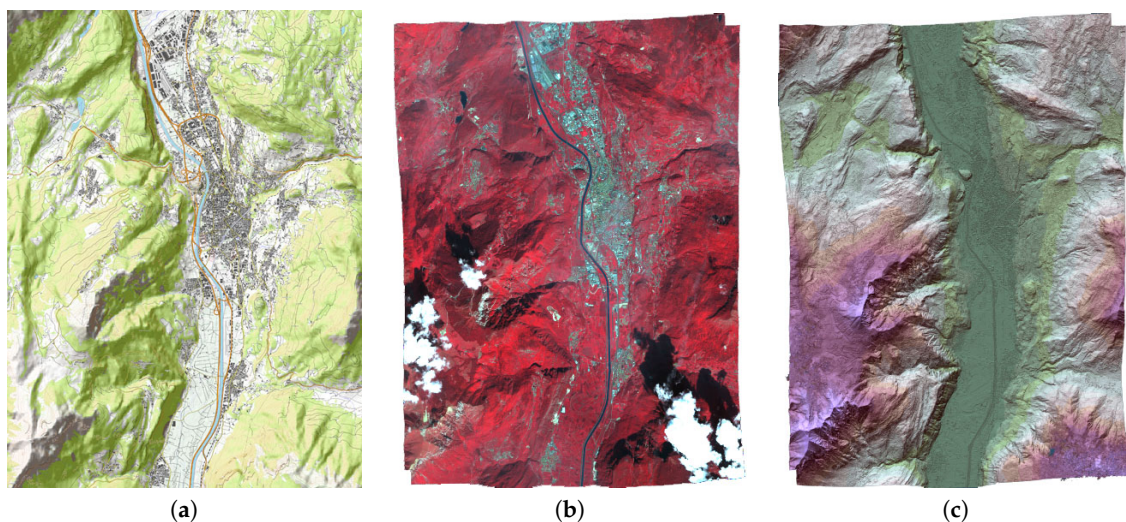


**Figure 4.** The test site Trento (Italy): (**a**) topographic map (opentopomap.org (CC-BY-SA)), (**b**) CIR ortho-image and (**c**) relief shaded DSM. (**b**,**c**) were produced employing the proposed and implemented mapping workflow.

### 3.3. Test Site Ljubljana (Slovenia)

This data set consists of three Pléiades stereo acquisitions from adjacent orbits over the region north to Ljubljana, Slovenia, first presented in Reference [5]. The stereo sets were acquired within three days in July 2013, the first and third set with the Pléiades-1A platform and the second with Pléiades-1B. The ellipsoidal terrain height of the region of interest ranges from 390 m to 1950 m and the scene covers about 400 km$^2$ consisting of agricultural land, managed forest, villages, the airport Brnik

and mountainous areas. The test site is shown in Figure 5. Reference data exists in terms of 18 GCPs and 12 ICPs measured in high resolution orthophotos, LiDAR reference DSM, LiDAR reference DTM and image coordinate measurements of the GCPs and ICPs. LiDAR reference data was taken in 2015 with a mean density of 14 points/m$^2$ over a region of 345 km$^2$. The LiDAR DSM and DTM with a GSD of 1 m were derived using OPALS [88]. This data set holds a temporal gap between LiDAR and Pléiades acquisitions of only two years. Thus, analysis can be performed on large regions as the structural changes are smaller than for the other data sets. For any other detail on the dataset we refer to Reference [5].

The acquisition scenario is depicted in Figure 6 and is of particular interest. In contrast to classical stereo or tri-stereo along track acquisitions, here also across track stereo pairs can be employed for 3D mapping, which is not the case for the other test sites and which was never done in literature so far. Regarding the multiple view geometry concept our data set holds more than the three along tracks stereo pairs from descending orbit, namely 15 stereo pairs. Table 2 lists the convergence angles of all possible pairs based on the equations in Equation (1). Four pairs have a small intersection angle of about 12.5° (1–3, 2–4, 3–5, 4–6), while all others have larger convergence angles from 23° to 38°.
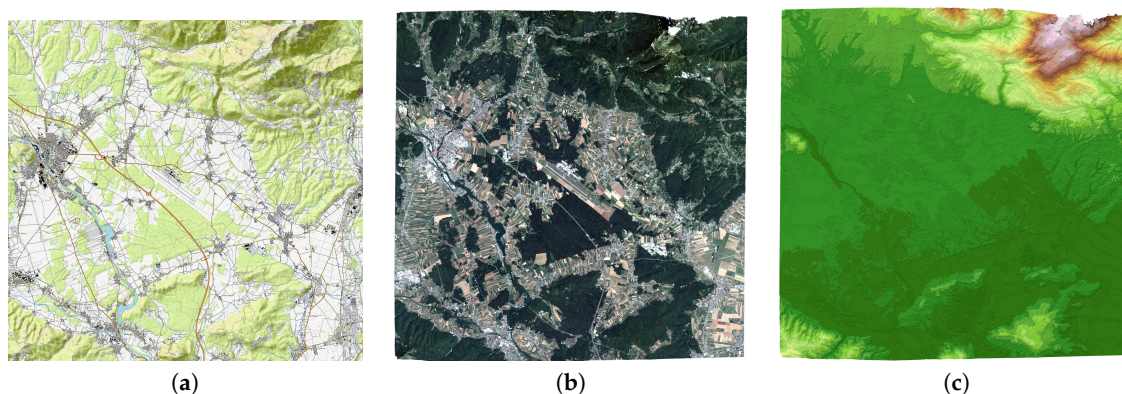
**Figure 5.** The test site Ljubljana (Slovenia): (**a**) topographic map (opentopomap.org (CC-BY-SA)), (**b**) RGB ortho-image and (**c**) relief shaded DSM. (**b**,**c**) were produced employing the proposed and implemented mapping workflow.

**Table 2.** Convergence angles for all possible stereo pairs of the Ljubljana data set given in degrees.

| ID | 2 | 3 | 4 | 5 | 6 |
|----|------|------|------|------|------|
| **1** | 25.2 | 12.2 | 27.5 | 24.0 | 33.9 |
| **2** |      | 26.4 | 12.5 | 37.6 | 25.0 |
| **3** |      |      | 22.7 | 13.0 | 24.7 |
| **4** |      |      |      | 30.3 | 12.6 |
| **5** |      |      |      |      | 26.2 |

### 3.4. Test Site Singapore (Singapore)

This test site covers a part of the island of Singapore, thus mainly consisting of a dense urban environment with very tall buildings. The height is ranging from sea level to about 200 m and the scene covers about 275 km$^2$. For Singapore terrestrially measured GCPs from Singapore Land Authority and the Tropical Marine Science Institute are available, which are used for sensor model validation and adjustment. However, due to restrictions the GCP heights are rounded to integer values and thus degrade the advantage of highly accurate in-situ measurements in this context. Of the whole set of control points, 16 GCPs are located within the specific stereo scene. Available LiDAR data is under restriction as well and cannot be used within this study.
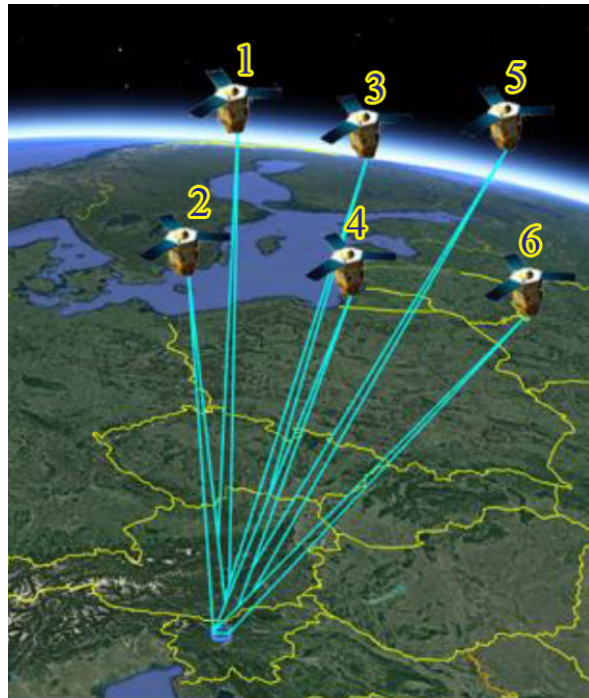
**Figure 6.** Multiple view geometry Ljubljana data set, adapted from Reference [5]. The Google Earth visualization depicts the preview of the footprints and the satellite's position.

### 3.5. Test Site Tian Shui (China)

The study area is located in central China close to the city of Tian Shui. Due to its location, it shows a geomorphological position in the landscape since it belongs to the Tibetan Plateau and also to the southern edge of Long-Zhong loess hilly area [110]. The ellipsoidal heights are ranging from 1200 m to 2100 m and the scene covers 395 km$^2$. Since for this site no reference data is available it more or less serves for visual interpretation purposes only. To allow accurate relative orientation tie-points are manually measured and applied within the sensor model adjustment. The test site is shown in Figure 7.
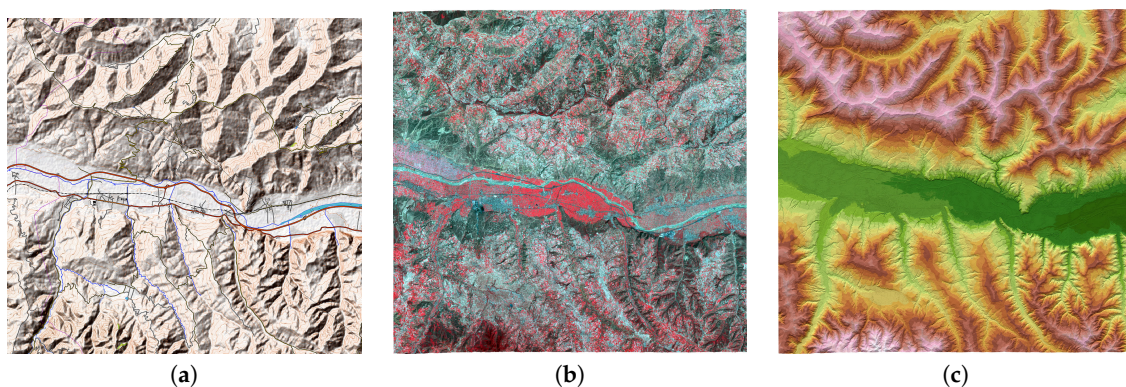


(**a**)                                          (**b**)                                          (**c**)

**Figure 7.** The test site Tian Shui (China): (**a**) topographic map (opentopomap.org (CC-BY-SA)), (**b**) CIR ortho-image and (**c**) relief shaded DSM. (**b**,**c**) were produced employing the proposed and implemented mapping workflow.

## 4. Methods

The envisaged end-to-end mapping workflow comprises four core tasks to be efficiently accomplished (cf. Figure 1). Thus, in this section we discuss our algorithmic approaches beyond the state-of-the-art with respect to (1) sensor modeling and parameter optimization, (2) generation of

DSMs from multi-stereo Pléiades images, (3) generation of DTMs using the extracted DSMs and (4) ortho-rectification of the multi-spectral image data. In addition, evaluation protocols are presented that are later employed in the assessment in Section 5. The structure of this section follows the flow graph Figure 2 and thus the related work presented in Section 2.

### 4.1. Sensor Modeling, Parameter Optimization and Geo-Location Accuracy Assessment

The RPC sensor model, as defined in References [40,41], relates geographic coordinates to corresponding image coordinates (i.e., a map-to-image transformation) by two rational cubic polynomials. The RPC model consists of 80 coefficients and approximates the physical sensor model in a standardized manner. For optimizing an RPC sensor model GCPs and their related image coordinates are employed. The optimization can be formulated in object-space or in image-space [41]. We pursue an adjustment in object-space since then the solution is also dependent on the height. This would not be the case if, for instance, an affine transformation is adjusted in image-space [40,42]. From the mathematical perspective, the parameter optimization can be formulated as a non-linear equation system. Like in Reference [41], we use Newton's method to solve the equation system and thus to optimize the sensor model.

To assess the geo-location accuracies, we follow the protocol that we defined in Reference [48]. The 2D geo-location accuracy of initial and optimized sensor models is evaluated for each images using reference GCPs and also their 2D locations in this image (cf. Figure 2). The 3D GCPs are projected onto the 2D image using the RPC model and the deviations to the reference 2D image coordinates define the model accuracy. Those differences are described using statistical measures such as mean, standard deviation (STD) or root mean square (RMS) errors in pixels. We give those values for across and along track direction such as the absolute length of the differences. Often those statistics are given for GCPs and ICPs.

For stereo acquisitions also the 3D geo-location accuracy can be determined. In this case, a spatial forward point intersection is performed based on the 2D image coordinates and the sensor models yielding 3D points. Those points are compared with the reference GCPs and ICPs resulting in statistics on 3D locations errors in meters (assuming that the underlying map projection is metric). Here the statistical values are given for East, North and Height direction such as for the absolute Euclidean distance (also called length).

### 4.2. Digital Surface Model Generation

As illustrated in Figures 1 and 2 the workflow for DSM generation comprises several steps that have to be applied sequentially and are described in detail in this section.

#### 4.2.1. Epipolar Rectification

For epipolar rectification of the Pléiades push-broom images our approach in Reference [54] is modified that initially was designed for SAR images. However, we show that it can also be applied to push-broom geometries by supporting the RPC sensor model within the framework. This aspect highlights this method since a single implementation is applicable to multiple geometries. For details of the rectification method we refer to Reference [54]. Since, the underlying transformations from image to epipolar rectified geometry are non-linear and also incorporate iterative image-to-map transforms (e.g., solved via Newton's method and linearization), pixel-by-pixel evaluation is computationally expensive. Especially, as the epipolar Pléiades images are huge, ca. $50,000 \times 50,000$ pixels for a full scene, an approximation is desired. It can be observed that the transformation is a locally slowly changing continuous function. Thus, we propose an algorithmic extension by a rigorous evaluation on grid points only. Transformation results between the grid points are then achieved by bilinear interpolation. We set the grid size such that the maximal pixel shift in column and line direction in comparison to the rigorous transformation is below 1/1000 of a pixel and thus negligible. For the test site Ljubljana, this accuracy is achieved when using a grid size of 250 pixels. We reduce it to

100 pixels to ensure errors below 1/1000 of a pixel also for any other possible scene, thus reducing the transformation effort by a factor of 10,000. Since the epipolar rectification maps an image from the original resolution to the same resolution, the GSD does not influence the mapping accuracy and thus the grid size does not need to be adjusted w.r.t. the GSD. Actually, the error limit of 1/1000 of a pixel can be seen as no error as we are working on the image domain. We could also set this limit to 1/100 of a pixel which then corresponds to a grid size of 700 pixels and thus a speedup of 500,000. To avoid recalculation of those grids they are also stored to be used later in the spatial point intersection. Those grids are called forward and backward addresses. Figure 8 depicts the geometric relation between stereo input images, their epipolar versions, forward and backward addresses and both disparity maps gathered in the matching step (cf. Section 4.2.2). For instance, a pixel coordinate from image 1 given in Figure 8 can be transformed into the epipolar version of image 1 by using the forward addresses of this image. Overall, the proposed approximation results in a significant speedup, thus extending the approach in Reference [54].
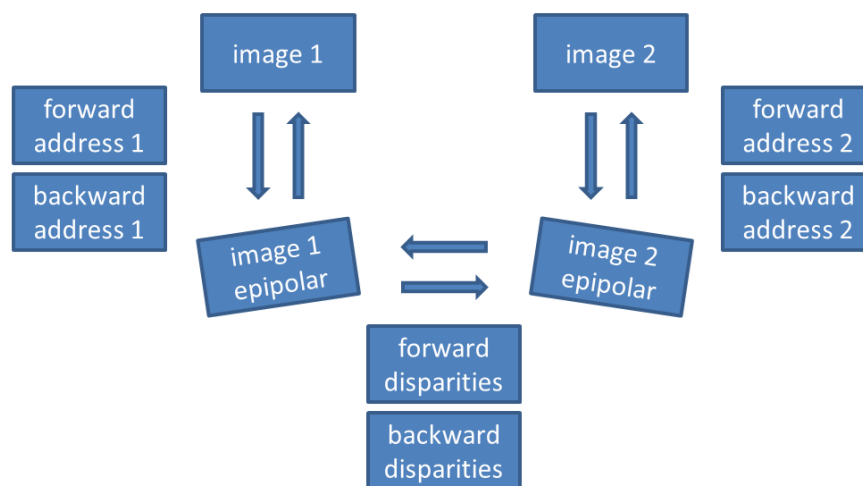


**Figure 8.** Relation of stereo input images, their epipolar rectified versions, forward and backward addresses and both disparity maps. Boxes represent images and arrows the geometric transformation between them.

### 4.2.2. Image Matching

As the image matching is employed on epipolar images, the disparities have to be searched along one dimension only, that is, the column direction. The classical matching paradigm is depicted in Figure 9 and consists of three main steps: disparity space image calculation (i.e., creation of the 3D cost volume), cost aggregation (i.e., an optimization by incorporating the neighborhood of a pixel) and disparity section (i.e., selection of the final disparity per pixel from the aggregated cost volume).

The current state-of-the-art in cost volume calculation is the normalized Hamming distance of the Census transform [72], as it is invariant to non-linear radiometric distortions and very efficient in its calculation. Since it compares all pixels within a patch to the central pixel value the transform becomes ambiguous when the patch holds homogeneous, that is, textureless, values. Therefore, the modified Census can be used that compares the pixels to the mean value of the patch [75,111]. We propose another option that works well for satellite-based images, which is to stabilize the Census transform by combination with a *sum of absolute differences* (SAD) cost function. SAD is normalized by the bit depth of the images, that is, $2^{16}$ for Pléiades. Normalized cross-correlation yields more blurring at 3D breaklines than Census transform such that it is not applied (cf. Reference [73]). It is also important to note that within this work the matching only uses the panchromatic band of the image. Obviously, the pansharpened multi-spectral information could be used as well. In this case, the combined cost-function is defined as the weighted sum of the costs for each spectral band. However, previous

experiments showed that color information is not helping much but slows down the calculation of the disparity image significantly [112]. The influence of the patch size on the performance in terms of runtime and accuracy is evaluated in Reference [73], where a $9 \times 9$ pixel patch performs best. In literature often $9 \times 7$ pixel patches are used as those 63 bits can be efficiently stored in a 64-bit register (cf. Reference [113]).
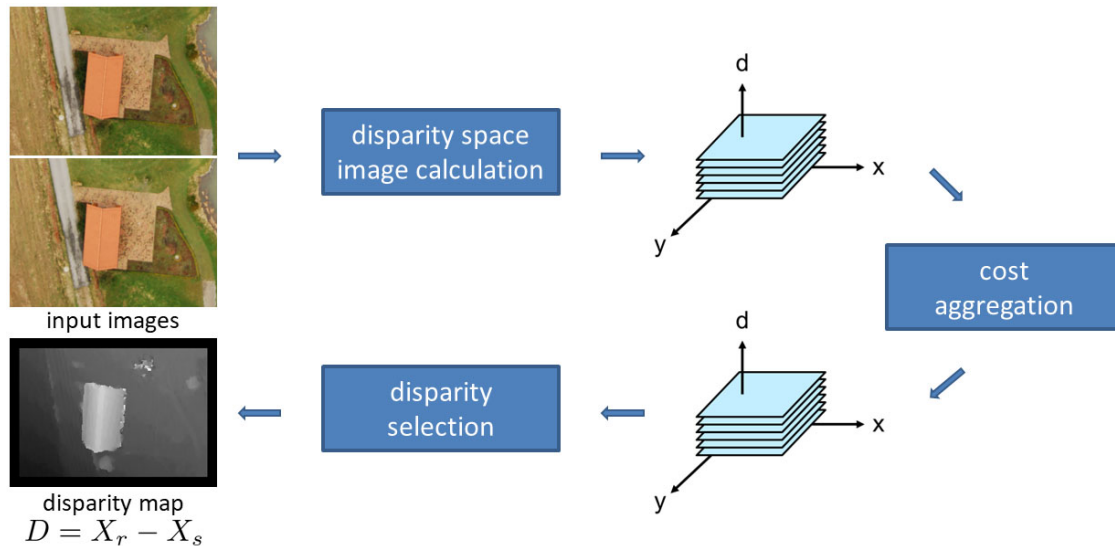


**Figure 9.** Image matching principle with epipolar rectified input images. The main steps are calculation of the disparity space image, cost aggregation and disparity selection. The final disparity map $D$ holds the shifts in column directions for each pixel such that they point from the reference image $X_r$ to the search image $X_s$.

After calculation of the cost volume the cost aggregation is performed with semi-global matching (SGM) introduced in Reference [61]. It combines the efficiency of a local method with an accuracy comparable to global methods. This is achieved by approximating a 2D Markov random field (MRF) optimization with several 1D scanline optimizations (cf. Reference [71]). Those scanline optimizations can be efficiently solved via dynamic programming yielding low runtimes. It has been shown that SGM is a special case of message passing algorithms like belief propagation and tree-reweighted message passing [69]. Overall, SGM approximates the energy minimization of a 2D Markov random field

$$E(D) = \sum_{p} C_{p}(d_{p}) + \sum_{p,q \in \mathcal{N}} V(d_{p}, d_{q}), \qquad (2)$$

with $C_{p}(d_{p})$ being a unary data term that represents the matching cost of a matching pixel $p$ at disparity $d \in \mathcal{D} = [d_{\min}, d_{\max}]$ and $V(d, d')$ is a pairwise smoothness term that penalizes disparity differences between neighboring pixels in the neighborhood $\mathcal{N}$. In SGM, $V$ implements a first-order smoothness constraint as

$$V(d, d') = \begin{cases} 0 & \text{if } d = d' \\ P_1 & \text{if } |d - d'| = 1 \\ P_2 & \text{if } |d - d'| \geq 2, \end{cases} \qquad (3)$$

with the penalty values $P_1$ and $P_2$. They penalize disparity jumps of 1 pixel and of multiple pixels. Now instead of minimizing the MRF, which is NP-hard, SGM minimizes a 1D version of Equation (2) along 8 cardinal directions $r$ via dynamic programming [61]. For each direction $r$ an aggregated matching cost $L_{r}(p, d)$ is recursively calculated as

$$L_{r}(p, d) = C_{p}(d) + \min_{d' \in \mathcal{D}} \left( L_{r}(p - r, d') + V(d, d') \right). \qquad (4)$$

Then, the 8 aggregated costs are summed at each pixel, yielding the aggregated cost volume

$$S(\boldsymbol{p}, d) = \sum_{\boldsymbol{r}} L_{\boldsymbol{r}}(\boldsymbol{p}, d) \,, \tag{5}$$

where the minimum at each pixel is chosen as the final disparity

$$d_{\boldsymbol{p}} = \arg\min_{d} S(\boldsymbol{p}, d) \,. \tag{6}$$

Since 3D breaklines are often visible in the input images as local changes of the grayvalues, the penalty $P_2$ can be adjusted w.r.t. these local changes of the reference input image $I$ in the current direction $\boldsymbol{r}$ [61]. As in the original paper no details on the adaptive penalty function are given, we propose to define $P_2$ as a function of $I$, $\boldsymbol{r}$ and three parameters $\epsilon$, $\tau$ and $P_2'$ as

$$P_2(I, \epsilon, \tau) = \max\left(\frac{\mu_I \cdot P_2'}{|I_{\boldsymbol{p}} - I_{\boldsymbol{q}}| + \epsilon}, P_1 + \tau\right) \tag{7}$$

with $\mu_I$ the mean gradient in column and line direction of $I$, $\boldsymbol{q}$ the adjacent pixel coordinate of $\boldsymbol{p}$ based on $\boldsymbol{r}$, $P_2'$ the initial penalty value and some $\epsilon$ to avoid a division by zero if $I_{\boldsymbol{p}}$ equals $I_{\boldsymbol{q}}$. The parameter $\epsilon$ is set to $\epsilon = \{\min(|I_{\boldsymbol{p}} - I_{\boldsymbol{q}}|) \mid \forall |I_{\boldsymbol{p}} - I_{\boldsymbol{q}}| \neq 0\}$, that is, $\epsilon$ becomes 1 for most images of type uint8 or uint16. The equation also ensures that $P_2$ is always larger than $P_1$ by at least a predefined $\tau$. We set $\tau$ in a pragmatic way to $\tau = P_1 + (P_2' - P_1)/100$ such that it automatically adapts to the input penalty values. This extension results in twofold improvements: (1) in homogeneous regions the penalty $P_2$ becomes huge such that large (mostly incorrect) disparity jumps are omitted and (2) at pixel locations with large local grayvalues changes $P_2$ become smaller and thus 3D jumps are cheaper for the optimization. Since 3D breaklines often correspond to grayvalue changes such jumps are better modeled by the algorithm.

The resulting disparities $d_{\boldsymbol{p}}$ are calculated at integer precision. To get subpixel refinements the classical approach is to fit a polynomial of order 2, that is, a parabola, into the three values surrounding the best-estimated disparity. The minimum of this polynomial, which can be extracted analytically, then defines the subpixel disparity.

For outlier detection and interpolation, the following options are implemented. A classical left-right consistency check is performed by employing backmatching. Disparities with a backmatching distance larger than a given threshold are marked as outliers. The standard threshold value is just above $\sqrt{2}$ such that a matching is correct if the backmatching hits within the local $3 \times 3$ window (e.g., 1.5 pixel). Then also a threshold on the aggregated cost is performed. Lastly, a peak filter detects connected regions (cf. Reference [61]), where small islands are discarded. In the next step, discarded pixels are classified as blunder or as occlusions as described in Reference [61]. Then, blunder locations are linearly interpolated based on their neighbors, while occluded locations are interpolated toward the ground level to avoid blurring at 3D breaklines.

Since the Pléiades images are huge, we propose an algorithmic extension mainly resulting in speed-up as follows. While in Reference [61] a hierarchical image pyramid based approach is used to determine the disparity search range for the final level (i.e., the original image), we apply the paradigm first presented in Reference [114] to limit the search range for each pyramid level (visualized in Figure 10). Instead of operating on the complete cost volume we propose to only use a band of this volume for disparity search. This change results in a significant speed-up, especially for scenes with large height variation, since we only have to search for a given number of disparities in the last level (e.g., 41 pixels instead of 2000 pixels, which is the average disparity range of the Ljubljana data set). Depending on the topography of the scene this cost volume (and also the aggregated cost volume) can be extremely large. For instance, on the Ljubljana test set this volume has a size of about 20 terabytes in 32 bit float encoding. With the proposed extension it shrinks to 2% of the original size, which is still

around 400 gigabytes. We call this method *truncated SGM*. As depicted in Figure 10 it might be the case that for a disparity $d$ the adjacent disparity $d'$ is not calculated thus does not exist. Therefore, we propose to modify the smoothness term given in Equation (3) to

$$
V(d, d') = \begin{cases} \text{if } d' \text{ exists} & \begin{cases} 0 & \text{if } d = d' \\ P_1 & \text{if } |d - d'| = 1 \\ P_2 & \text{if } |d - d'| \geq 2 \end{cases} \\ \text{else} & P_2 \, . \end{cases} \tag{8}
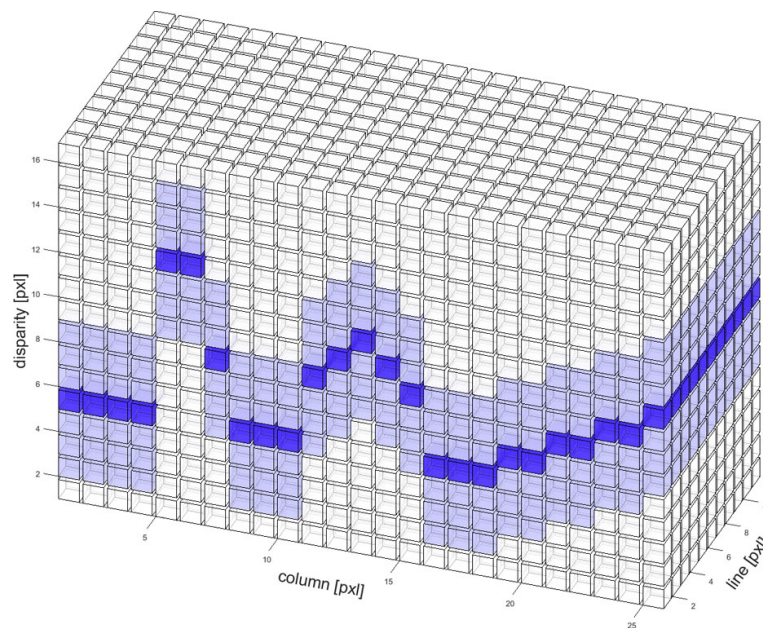$$



**Figure 10.** Cost volume $C_p(d)$ of classical SGM versus our implementation is shown on a toy example. While SGM evaluates the whole cost volume, we only check the disparities visualized in light blue color around the disparity predictions from the previous pyramid level depicted in dark blue color. In this toy example the search space is reduced to 44%, however, the reduction is larger in real satellite images.

Depending on the 3D structure of the scene maximal memory consumption reduction between 65% to 99% is achieved while the runtime is reduced by 30% to 98% (also cf. Reference [65]). In addition, instead of storing the cost volume and the aggregate cost volume as 32-bit float values, we discretize them to 16-bit unsigned integers (uint16). This is done by introducing a scaling factor and yields another 50% memory reduction. This discretization obviously affects the results. However, the rationale is that the discretization changes the original values only a little bit (some fractions as the initial costs are in the range $[0, 1]$). Thus, if those small changes influence the argument of the minimum of the aggregated costs given in Equation (6), then those disparity values are very likely incorrect anyway. Such regions are later detected in blunder and occlusion classification and removed. Overall, the discretization has a non-significant influence on the final disparity map. Next, we observe that due to the prior epipolar rectification a significant region on the border of the input images consists of nodata values. In cases where all disparities of investigation only contain such invalid regions, the SGM optimization could be skipped. Therefore, we introduce a specific uint16 value for the disparity space image that indicates those regions. The runtime of the SGM optimization can then be reduced by around 33% (mean value over our test images).

To gain speedup and accuracy, we incorporate a coarse DSM (e.g., SRTM DEM [115], ASTER GDEM [116], World DEM [117]) on the top pyramid level to calculate disparity predictions and thus to limit the search space and reduce ambiguities. For a grid of pixels in the reference image,

an image-to-map transformation is calculated using the coarse DSM. The resulting map coordinates are then projected into the search image with a map-to-image transformation. The couple of 2D image coordinates is then transformed into the epipolar images using the forward addresses (cf. Figure 8). The column difference of the epipolar pixel coordinates then serves as prediction for image matching. With such disparity prediction the search space can be significantly reduced at the top level.

Like others, we found that the simple parabola subpixel interpolation yields a bias in the distribution of fractional disparities (cf. Reference [118]). This bias causes incorrect reconstructions of tilted surfaces. To reverse this effect we calculate a mapping between the current fractional distribution and a uniform distribution. It is then applied efficiently via a lookup table. This effect is depicted in Figure 11.
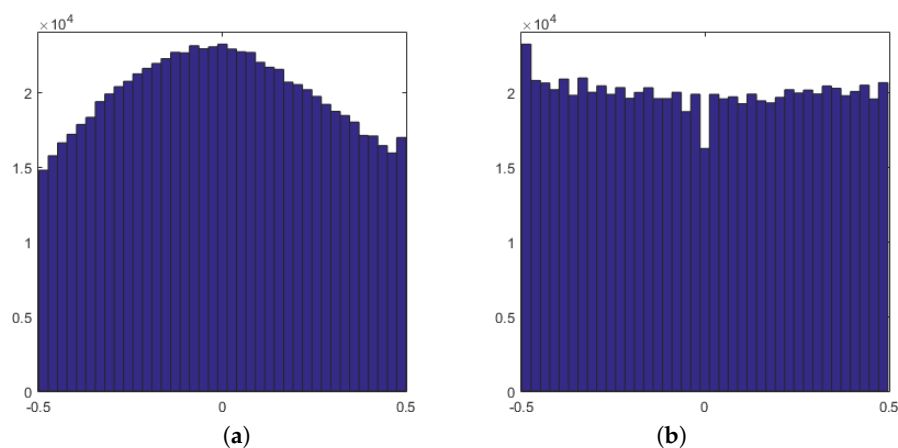


**Figure 11.** Subpixel distribution (**a**) based on the parabola interpolation [61] and (**b**) after our proposed normalization.

### 4.2.3. Spatial Point Intersection

From each valid matching result (i.e., disparity) a 3D point can be calculated by spatial point intersection, finally yielding a 3D point cloud irregularly distributed in space. In the first step, for each pixel in the disparity map, the locations in the reference and search input images have to be determined. As indicated in Figure 8 the backward address of the reference image and the backward address of the search image are employed in this step. Then, the spatial point intersection can be performed via an iterative least squares adjustment. The resulting point cloud can be stored in LAS or LAZ format [119] or in a multi-band image. In this paper, the second option is used, where a three-band image of the size of the disparity map then stores East, North and height values (cf. Figure 12a and [84]).

### 4.2.4. DSM Resampling

Since in remote sensing a classical data representation is still the 2.5D DSM the 3D point cloud has to be mapped to a 2D image with equidistant spacing (i.e., a regular grid). This process is also called resampling or regridding. We present two simple approaches to doing so. First, the dimension and spacing of the output DSM are defined. Then, we use either nearest neighbor interpolation or areal interpolation in the regridding step. In the former case, each 3D point is projected into the DSM and put into the nearest integer position. If this value was already assigned then the maximum between the previous and the current value is chosen. In this way, the height value closest to the sensor is reconstructed. In areal interpolation, four adjacent 3D points are projected into the DSM. Then, all integer locations that are covered by this quadrangle are linearly weighted interpolated. Double assigned pixels are treated as described above. Overall, this procedure is a direct resampling that potentially is not hitting all pixels of the DSM grid. However, as multiple DSMs get fused in further processing, this simple strategy is sufficient and in reality the nearest neighbor approach is usually used due to performance benefit. The DSM resampling is depicted in Figure 12.
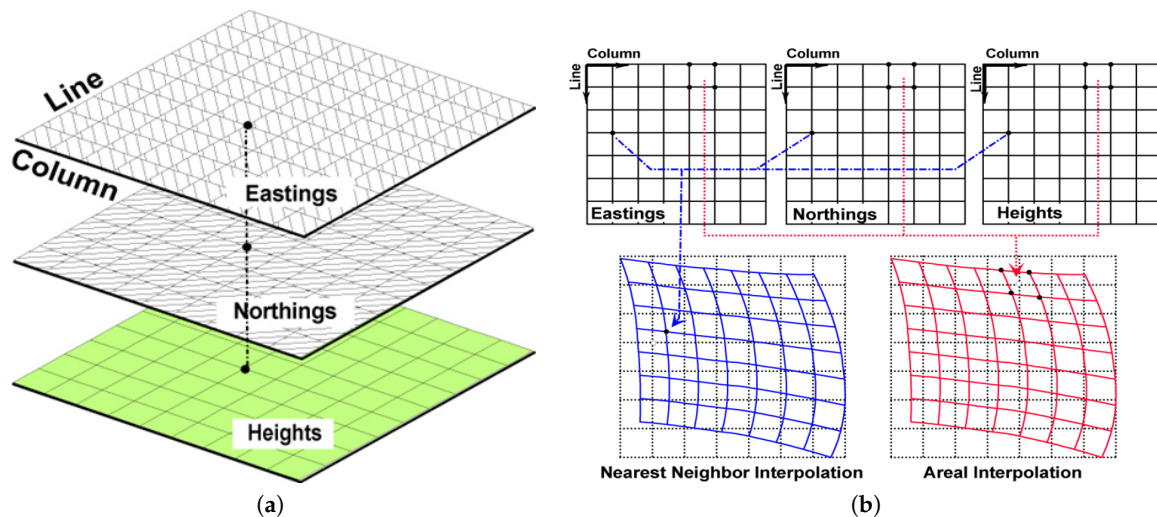
**Figure 12.** DSM resampling: (**a**) representation of 3D point cloud as a multi-band image and (**b**) regridding options.

### 4.2.5. DSM Fusion

In the proposed processing chain, the DSM fusion is applied for two main reasons: (1) to combine the information of forward and backward matching in cases when only a single stereo input set exists and (2) to combine multiple forward and backward matchings in cases when multiple input (stereo) sets exist. Thus, the fusion process gets multiple DSMs as input and calculates one final DSM. For simplicity, we stick to a local method [85] which takes all height measurements within a $3 \times 3$ pixel neighborhood and extracts the probability mode of this probabilistic height distribution. The only required parameter is a threshold value of how many input height values have to exist such that an output is generated. To get high coverage this value is set to 1 by default. We extend this method for arbitrary neighborhoods and also incorporate a step size. This step size can be used to generate a lower resolution DSM by, for instance, only extracting the probability mode at every second pixel in column and line direction. In contrast to a straightforward median-based approach, local height errors can better be eliminated using the probability mode [85].

### 4.2.6. DSM Assessment

In remote sensing, it is common sense to compare two DSMs based on their differences. This leads to a distribution defined as $\text{DSM}_\Delta = \text{DSM}_{\text{ref}} - \text{DSM}_{\text{current}}$, where in our case the reference is a LiDAR DSM and the current DSM the one retrieved from Pléiades image data. For judging the quality of the differences the statistical values mean and standard deviation are used. However, this only makes sense if the underlying distribution is a normal distribution. As observed by researchers this is actually not the case (cf. References [42,120–122]). First, both DSMs include strong outliers such that the distribution of differences contains heavy tails. Second, the distribution itself is often non-Gaussian as, for instance, gaps between buildings are not reconstructed from satellite data causing a systematic bias next to the main lobe. In addition, the mean and standard deviation measures are non-robust estimates which suffer from both aspects. Therefore, in this case, robust estimates have to be used; in particular, the median (MED) value for the center of our normal distribution and the normalized median absolute deviation (NMAD) for the standard deviation. The NMAD is defined as as

$$\text{NMAD}(\Delta H) = 1.4826 \cdot \text{median}(|\Delta H - \text{median} \, \Delta H|). \tag{9}$$

Figure 13 depicts a real example of this study where only errors between $-10$ m and $+10$ m are plotted. In particular, we show the distributions of height differences of a stereo DSM from test site Ljubljana (blue) with the classical normal distribution fits based on mean and standard deviation (red)

and the robust fits based on median and nmAD (green). From this plot, it is clearly visible that the standard deviation of the classical fit is far too large which is a result of the non-robust estimate due to outliers. The NMAD based normal distribution is narrower and fits the input data better. The estimated central value of the default fit is drawn to the negative side. This obviously happens with the robust fit as well, however, less distinctive. This aspect has to be kept in mind when interpreting height differences and thus, will later be discussed in Section 5.
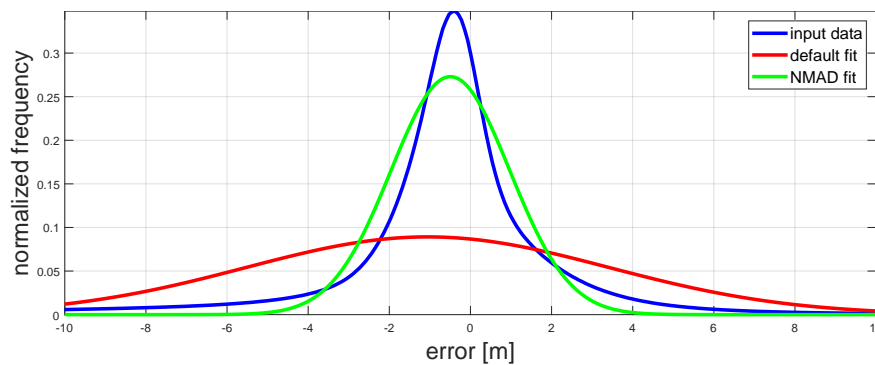


**Figure 13.** Example of fitting a normal distribution into non-normal distributed input data (blue). Default fit based on mean and standard deviation estimates (red) and robust fit based on median and nmAD estimates (green).

### 4.3. Digital Terrain Model Generation

To extract a terrain model from a surface model, man-made structures and vegetation have to be removed from the given input DSM. As depicted in Figure 1, the workflow for DTM generation comprises two main steps: determination of bare-earth points and interpolation of all other points and according post-processing. As no state-of-the-art method was found that delivered satisfying DTMs at fast runtimes, a novel method named *multi-directional slope dependent (MSD)* DTM generation is presented in this section. Our Matlab source code is available at GitHub: https://github.com/rolandperko/dsm2dtm.

#### 4.3.1. MSD DTM Generation Approach

The presented DTM extraction approach is especially suited for DSMs that are generated by a state-of-the-art photogrammetric workflow from VHR satellite stereo or tri-stereo images. One constraint of such DSMs is that 3D breaklines are not always clearly defined, which can be traced back to occluded areas that cannot be reconstructed. Our idea is to extend the algorithm of Reference [93], which is designed for LiDAR data, to be slope dependent and to be multi-directional, that is, 8-directional to span the complete 2D image space. Nonetheless, the focus is put onto simplicity, robustness and computational efficiency to follow integration needs into the automatic end-to-end workflow:

- Determine points in the DSM that are located in bare-earth regions.
- Apply some post-processing on the ground mask and remove all other non-bare-earth regions.
- Fill the resulting holes using DTM interpolation.
- Apply some post-processing.

Obviously, the crucial and most difficult step is the first one. Based on our literature study the most promising algorithm to start with is the directional filtering method by References [91,93]. A short recap on directional filtering according to Reference [93] can be given as follows: The basic idea is to process each line of the given DSM separately, for example, from left to right with a sliding window of a given extension. First, the minimal value in this window is determined, which is considered to represent the bare-earth terrain at the minimal position. In this step, it becomes obvious that an object

to be filtered has to be smaller than the filter extent. Then, if the current pixel under examination has a large difference to the minimal value w.r.t. a given height threshold, it is considered as a non-ground point. If this is not the case and the slope between the current pixel and the next one in scanline direction is larger than a given slope threshold, the pixel is also considered as a non-ground pixel. If the slope is positive and smaller than this threshold the pixel gets the same label as the previous pixel. If the slope is negative, then the distance to the closest ground point is used to decide whether the pixel is classified as ground or as non-ground. This method suffers from two drawbacks:

1. A bottleneck is inherent to the negligence of the local slope of the underlying terrain, such that no useful results can be expected on tilted surfaces or in mountainous areas.
2. Since it works on 1D image profiles, hereby denoted as scanlines, it yields fast runtimes. However, when applying single scanline processing only, the complete 2D context into which an object is embedded is lost. In addition, the given DSM can have unsharp 3D edges in some scanline directions.

Using these simple insights, two main extensions to the directional filtering concept are proposed for the development and implementation of our novel MSD DTM generation method:

1. Incorporation of the local terrain slope.
2. Extension of the local horizontal scanline approach to multiple scanlines, which spans the full 2D image space.

The main reason for the simplicity of the MSD algorithm is the fact that the robust fitting is replaced by a smoothing done beforehand. Now the terrain slope fit is directly gathered using the difference value of this smoothed input DSM. After removing the non-ground points, the resulting holes are filled employing a triangulation based linear interpolation. The algorithm uses three parameters, that is, the filter extent in meters, the height threshold in meters and the slope threshold in degrees.

### 4.3.2. Consideration of Local Slope

The main drawback in References [91,93] is that the local slope of the terrain is not considered in the processing. If this algorithm is used to process regions which are not flat, incorrect filtering occurs since the difference of the current height to the estimated minimal value depends on the terrain slope. If ignored the following height thresholding also removes bare-earth regions. This issue is sketched in Figure 14, where a 1D profile of an artificial DSM to be filtered is given in (a). It shows a tilted surface and a building with some noise. The potential filter extent is visualized as the blue dashed line. The minimal height value within the filter extent is marked as black dot. It is always dragged towards the terrain fall and thus is incorrect (in the example a height of 93.9 m). If we apply a robust terrain slope fitting, the green dashed line in (b) is received. Using that tilt, the initial profile could be slope corrected (c) and the new minimal value can now be correctly extracted (98.7 m in this example).

In addition to such an incorrect minimal value, the slope estimate of two adjacent DSM values should be corrected w.r.t. the local terrain slope as well.

The first algorithmic extension is very intuitive and sketched in Figure 14. In the workflow, a robust fit of the height values within the filter extent must be performed. An obvious solution would be to use an iterative weighted least squares method with bisquare (Tukey) or Huber weighting function (cf. Reference [123]). However, this robust fit has to be performed for each pixel and for all scanline directions and thus would slow down the process. Therefore, we propose a heuristic which yields similar results as the robust fit but can be calculated very efficiently. The idea is to smooth the DSM using a huge Gaussian kernel. Local objects like buildings disappear while the terrain slope remains in this smoothed surface. The slope heuristic in the desired direction can then directly be calculated as the pixel difference of two adjacent pixels.
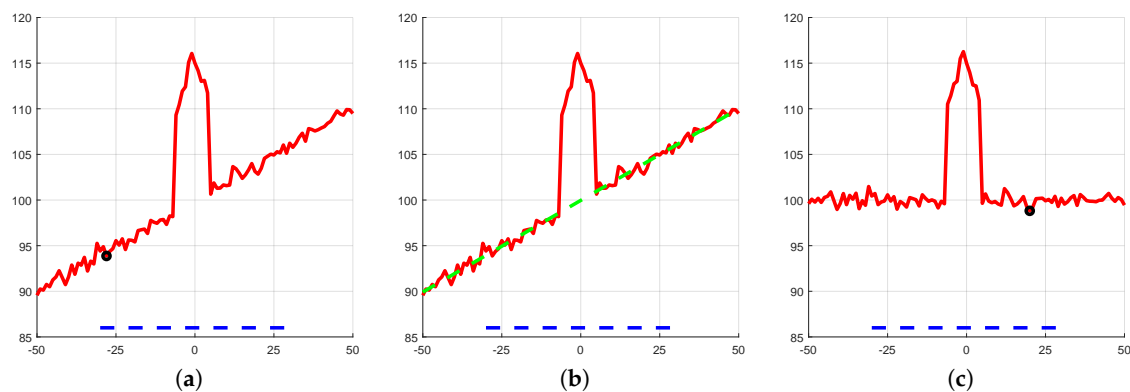
**Figure 14.** DSM profiles for an artificial building on a tilted surface with added noise: (**a**) Original DSM profile and the minimal height value (black) within the window (blue). (**b**) Robust fit of the surface (green). (**c**) Slope corrected DSM profile and the according minimal height value.

Therefore, a 2D Gaussian smoothing is implemented with a spatial sigma $\sigma$ and a kernel size of $n \times n$ m$^2$. Per default those parameters are set to $\sigma = 25$ m and $n = 101$ m. The (separable) 2D convolution is implemented as two 1D convolutions for speedup. The pixel difference of the central (smoothed) pixels then defines the local terrain slope value. On the Innsbruck East triplet data set the differences of the robust and heuristic slope estimates have a mean value of 0.05° with a standard deviation of 1.76°. These differences are small and acceptable, especially taking the speedup into account.

Figure 15 shows a real example for the downward scanline. Given are a subset of the DSM and the local slope estimate as achieved by smoothing. The plot shows the original 1D DSM profile (black), the robust fit of the terrain slope (green), the proposed simple fit of the terrain slope (cyan) and the robust and simply corrected data (red and blue). It is obvious that the simple and fast Gaussian-based approximation yields very similar results in comparison to the iterative bisquare weighted least squares solution. Actually, initial tests revealed that it is faster by a factor of at least four magnitudes for a filter size of 91 pixels (i.e., a speedup factor of at least 10,000).

### 4.3.3. Multi-Directional Filtering Approach

The second algorithmic extension concerns the data processing scheme in terms of scanlines. As stated above, consideration of only one scanline direction from left to right yields a local solution and the filtering result could abruptly change within two adjacent pixels from two neighboring scanlines. Instead, we propose to use 8 scanlines and fuse the results for final pixel classification. The principal concept to solve a problem in 2D by fusing multiple 1D solutions is based on the work of Reference [61]. In the presented case the fusion is simply based on majority voting. If more than 5 scanlines classify the current point as a ground point, this point is classified as ground point, else as non-ground point. This aspect is very important for satellite stereo DSMs, as some scanlines may classify a non-ground point incorrectly (e.g., due to a smooth height transition) while the combination of classifications achieved from 8 directions certainly helps to improve final pixel classification accuracy.

The complete novel algorithm is described in Algorithm 1 resulting in a ground and a non-ground label image. Instead of processing each scanline sequentially, the image can be scanned in two passes, first from the upper-left to the lower-right corner and second from the lower-right to the upper-left corner, processing four scanlines at each pass (cf. Figure 16). Doing so, image data has to be read only once and instead of storing eight label images, one label image can be used to sum up all ground pixels, followed by thresholding as stated above. This specific processing reduces the memory consumption by a factor of 8 and results in a minor speed-up since the input data is only read once.

(**a**) DSM  (**b**) local downward slope



(**c**)
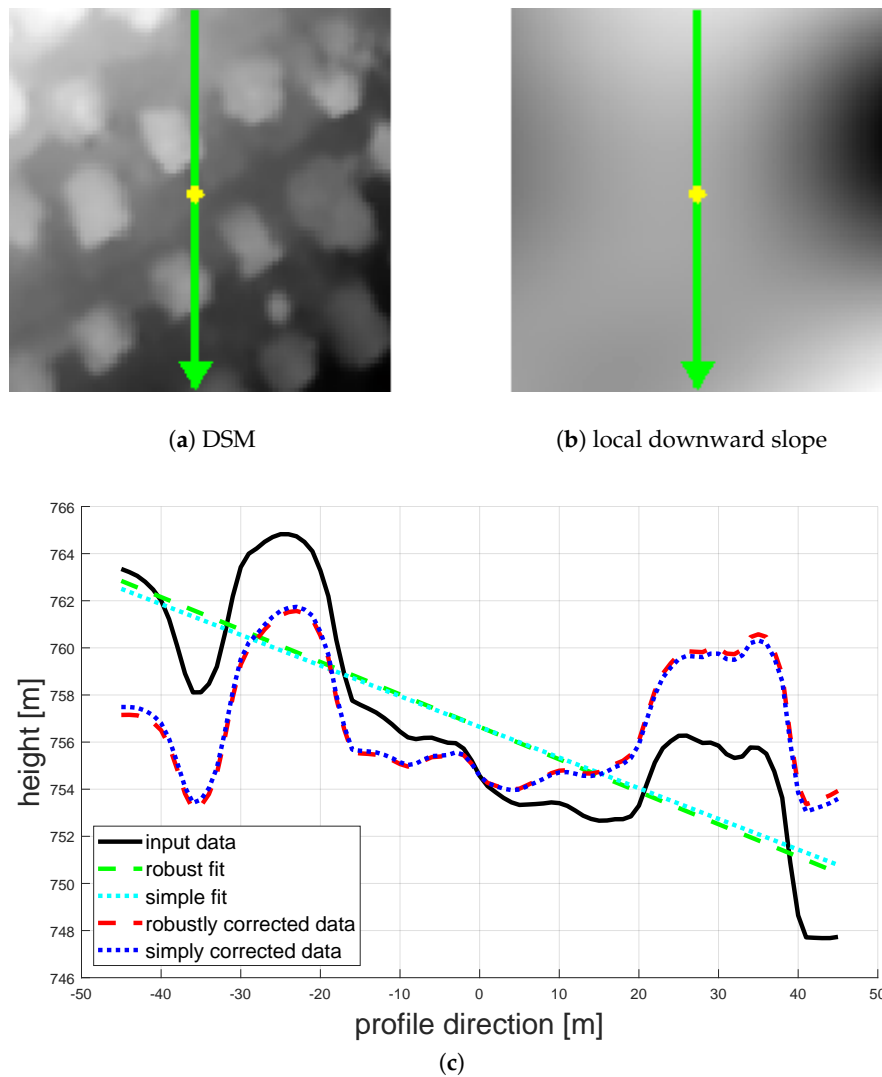
**Figure 15.** Comparison of robust terrain slope fit to the proposed simple fit. (**a**) shows a subset of a DSM and (**b**) its Gaussian filtered variant. (**c**) visualizes the profile of the DSM along the green arrow together with the terrain slope fits and the corrected profiles.
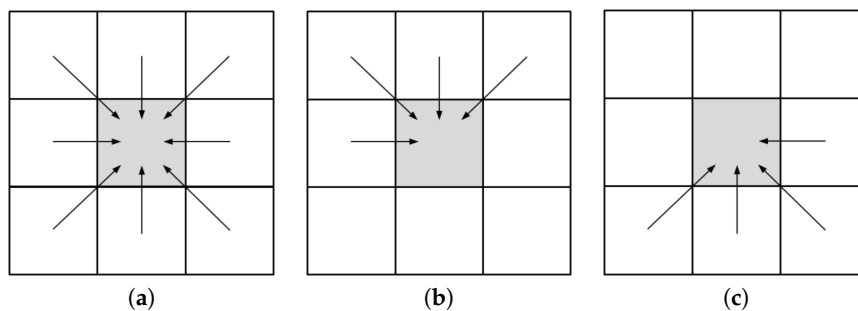


(**a**)  (**b**)  (**c**)

**Figure 16.** Concept of splitting 8 scanlines spanning the 2D image space (**a**) into 2 passes (**b**) down-right and (**c**) up-left.

---

**Algorithm 1:** Multi-directional slope dependent DTM generation.

---

    **Input:**

1        DSM                  `// Digital surface model.`

2        Spacing        `// Spacing of DSM in meters.`

3        DSMsmooth        `// Smoothed DSM.`

4        Dir     `// Scanline direction represented as two integer shifts ShiftX and ShiftY.`

5        ExtendMetric        `// Filter extend in meters,`
                      `// converted to pixels using Spacing and yielding Ext.`

6        Ext        `// Filter extend (odd) which also defines the local`
                      `// direction dependent neighborhood ExtX(Dir) and ExtY(Dir).`

7        X        `// centered number line with` $X = [-(Ext-1)/2 : +(Ext-1)/2]$.

8        ThrHeightDiff        `// Height difference threshold in meters.`

9        ThrSlope        `// Slope threshold in degrees.`

   **Output:**

10       Label $\in$ [Ground,NonGround]        `// Ground or non-ground classification.`

11  **Function** `DSM2DTM`(DSM, Spacing, DSMsmooth, Dir, ExtendMetric, ThrHeightDiff, ThrSlope)**:**

12    **foreach** ( scanline s ) **do**        `// Loop over whole image in x and y direction.`

13

14        DSMDiff = DSM(x,y) $-$ DSM(x+ShiftX,y+ShiftY)        `// Local height difference.`

15        DSMsmoothDiff = DSMsmooth(x,y) $-$ DSMsmooth(x+ShiftX,y+ShiftY)    `// Local`
                      `// terrain slope.`

16        Neigh = DSM(x+ExtX(Dir),y+ExtY(Dir))        `// Neighborhood (proposed filter extend).`

17        NeighCorr = Neigh + X*DSMsmoothDiff(x,y)        `// Slope corrected height values.`

18        MinNeigh = min(NeighCorr)        `// Slope corrected minimal terrain value.`

19        HeightDiff = DSM(x,y) $-$ MinNeigh        `// Difference to minimum.`

20        **if** ( HeightDiff $>$ ThrHeightDiff ) **then**

21          Label(x,y,s) = NonGround        `// Pixel is non-ground.`

22        **else**

23          Delta = DSMDiff $-$ DSMsmoothDiff        `// Slope corrected height difference.`

24          SignDelta = -sign(Delta)        `// Negated sign of Delta.`

25          SlopeLocal = atan2(abs(Delta),Spacing)*180/pi        `// Local slope.`

26          Slope = SlopeLocal*SignDelta        `// Corrected slope.`

27          **if** ( Slope $>$ ThrSlope ) **then**

28            Label(x,y,s) = NonGround        `// Pixel is non-ground.`

29          **else**

30            Label(x,y,s) = Label(x-ShiftX,y-ShiftY)        `// Assign as last label.`

31          **end**

32          **if** ( Slope $<$ 0 ) **then**

33            Label(x,y,s) = Ground        `// Pixel is ground.`

34          **end**

35        **end**

36    **end**

37    **return** Label        `// Return classification.`

---

4.3.4. Extension

The following extensions could be easily performed, where currently only the first one is supported by our implementation. The second and third options are very generic and independent of the proposed MSD DTM generation method, such that they are not evaluated here.

- Apply a peak filter after the filtering step to get rid of small, mostly incorrect, regions. This idea is similar to the filtering of disparity maps in image matching.
- Since, taking the minimal value of a distribution is not robust to outliers, the *n*th percentile could be taken instead. Then the algorithm becomes robust toward outliers in the input DSM.
- Usage of more appropriate hole filling techniques by, for instance, using Kringing [124], anisotropic diffusion [125], total variation [98] or total generalized variation [126].

*4.4. Ortho-Rectification*

Standard direct ortho-rectification is performed for each image. In the presented workflow, the rectification is embedded in the DSM resampling step during DSM generation. There, the multi-spectral images are available as epipolar rectified products. In the step of spatial point intersection, those points get their 3D coordinates and their multi-spectral values. Thus, the DSM resampling step is also applied to generate the ortho-images. For mosaicing, the ortho-images are sorted w.r.t. to their global incidence angles such that the more nadir ones are prioritized. During fusion, a radiometric block adjustment is performed based on a least squares adjustment.

## 5. Results

This section reports on a multitude of results based upon the previously presented methodologies. In particular, sensor modeling and optimization, digital surface model generation and digital terrain model derivation are addressed. First, the 2D and 3D geo-location accuracy is assessed for all images and stereo sets. This investigation includes a study on different parameters that are adjusted in the sensor modeling step. Second, the quality of the resulting DSMs is assessed. Here an investigation of the epipolar geometry, qualitative and quantitative evaluation of DSM accuracy w.r.t. reference LiDAR data and stereo versus multi-view stereo experiments are given and discussed. Third, DTM extraction is performed and results are compared to LiDAR reference data.

*5.1. Sensor Modeling and Parameter Optimization*

As first task, the 2D geo-location accuracy of all sensor models is evaluated following the assessment protocol defined in Section 4.1. The resulting statistics are presented in Table 3. Statistics are given in across and along track direction such as for the absolute 2D length. Here, it can be observed that most of the images have mean circular errors (represented as mean initial length) below 17 pixels, which corresponds to the 8.5 m CE90 as reported in Reference [32]. The Trento triplet holds the largest geo-location errors even above the CE90 values.

As second task, all sensor models are adjusted by optimizing the constant and linear RPC terms and the resulting RMS values are also reported in Table 3. The across and along track residuals are all in the sub-pixel range (0.3 to 0.9 pixels) and show a widely homogeneous behavior. According to the nominal GSD of 0.5 m, these pixel values correspond to geo-location errors in the range of 0.15 m to 0.45 m on ground. As discussed in Section 3, the lower bound of those errors is the accuracy of the reference data. As the reference was manually measured for some test sites better results cannot be expected. Note, that the accuracies are rather high when considering the reference LiDAR data with 1 m GSD for some of the test sites. Since the discretization and manual measurement errors are assumed to be of Gaussian distribution, the least squares adjustment may find solutions which are better than the individual measurements.

As third task, a specific test is performed on the image 1 of the Innsbruck stereo test data (cf. Table 1 and Figure 3). The 2D geo-location accuracy of this selected sensor model is evaluated where the parameter optimization is once based on releasing only the constant term of the RPC nominators and once by releasing the constant and linear terms. To doing so, the given 30 GCPs are divided into different sets of GCPs and ICPs. The results of this analysis, including mean and STD of checkpoint residuals, are summarized in Table 4. The table shows that utilization of the absolute minimum number of GCPs (i.e., 1 for constant and 4 for constant and linear coefficient optimization) yields systematic

errors for both scenarios, as expressed by the mean residual values. Appropriate over-determination, for example, utilization of 10 GCPs in this assessment, reduces such systematic errors significantly. The numbers also show that removing the shift only might be sufficient to achieve reliable accuracy, which is also confirmed in Reference [42]. Overall, we propose to perform the constant and linear-based adjustment since smaller residuals are retrieved and potentially systematic errors can be reduced.

As fourth task, another specific test is performed on all six images from the Ljubljana test site. Here, the sensor models are adjusted based on the optimization of constant and linear terms with the given GCPs. Then, those models are also evaluated on the ICPs to assess if an over-fitting occurs. Statistics are presented in Table 5. Here, the mean standard deviations after adjustment are 0.47 pixels for GCPs and 0.73 pixels for ICPs. As expected, the models adjust to the GCPs but also give decently small errors on ICPs.

Overall, all models show initial displacements and thus have to be corrected before continuing with the mapping workflow.

**Table 3.** 2D geo-location accuracy of Pléiades image data given in pixels. Shown are the initial (a-priori) mean and RMS residual errors and the RMS after parameter adjustment.

| Study Area | ID | GCPs | Mean Initial [pxl] | | | RMS Initial [pxl] | | | RMS Adjusted [pxl] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Across | Along | Length | Across | Along | Length | Across | Along | Length |
| Innsbruck | 1 | 30 | −8.55 | −5.91 | 10.43 | 8.61 | 5.96 | 10.47 | 0.80 | 0.50 | 0.95 |
| Stereo | 2 | 30 | −8.55 | −2.28 | 8.88 | 8.61 | 2.42 | 8.94 | 0.78 | 0.61 | 1.00 |
| Innsbruck | 1 | 21 | −9.61 | −8.55 | 12.88 | 9.63 | 8.60 | 12.91 | 0.57 | 0.80 | 0.99 |
| Triplet | 2 | 20 | −10.21 | −9.49 | 13.96 | 10.24 | 9.54 | 14.00 | 0.76 | 0.82 | 1.12 |
| East | 3 | 20 | −10.14 | −11.50 | 15.35 | 10.16 | 11.53 | 15.37 | 0.53 | 0.79 | 0.96 |
| Innsbruck | 1 | 16 | −8.93 | 4.43 | 9.99 | 8.97 | 4.47 | 10.02 | 0.62 | 0.39 | 0.73 |
| Triplet | 2 | 16 | −8.93 | −0.18 | 8.97 | 8.95 | 0.91 | 8.99 | 0.44 | 0.65 | 0.78 |
| Center | 3 | 16 | −9.10 | 4.18 | 10.03 | 9.13 | 4.22 | 10.06 | 0.58 | 0.32 | 0.66 |
| Innsbruck | 1 | 22 | −9.62 | −8.56 | 12.92 | 9.67 | 8.59 | 12.94 | 0.46 | 0.60 | 0.75 |
| Triplet | 2 | 22 | −9.23 | −11.46 | 14.73 | 9.27 | 11.49 | 14.76 | 0.66 | 0.62 | 0.91 |
| West | 3 | 22 | −9.82 | −9.53 | 13.70 | 9.85 | 9.55 | 13.72 | 0.48 | 0.61 | 0.78 |
| Trento | 1 | 21 | 11.19 | 12.59 | 16.90 | 11.22 | 12.70 | 16.95 | 0.75 | 0.76 | 1.07 |
| Triplet | 2 | 21 | 10.80 | 13.81 | 17.58 | 10.84 | 13.91 | 17.63 | 0.67 | 0.71 | 0.98 |
| | 3 | 21 | 8.61 | 15.71 | 17.97 | 8.65 | 15.81 | 18.02 | 0.81 | 0.88 | 1.20 |
| Ljubljana | 1 | 28 | 1.63 | −0.05 | 1.83 | 1.81 | 0.65 | 1.93 | 0.55 | 0.60 | 0.82 |
| Stereo 1 | 2 | 28 | 2.70 | 2.51 | 3.79 | 2.89 | 2.60 | 3.89 | 0.65 | 0.58 | 0.87 |
| Ljubljana | 3 | 30 | 1.41 | −0.63 | 1.74 | 1.61 | 0.85 | 1.82 | 0.54 | 0.52 | 0.75 |
| Stereo 2 | 4 | 30 | 1.42 | 1.72 | 2.44 | 1.81 | 1.88 | 2.61 | 0.66 | 0.71 | 0.97 |
| Ljubljana | 5 | 30 | −3.07 | 11.66 | 12.07 | 3.13 | 11.67 | 12.08 | 0.63 | 0.60 | 0.87 |
| Stereo 3 | 6 | 30 | −2.93 | 12.14 | 12.50 | 2.99 | 12.16 | 12.52 | 0.54 | 0.73 | 0.91 |
| Singapore | 1 | 14 | 4.42 | 1.03 | 4.64 | 4.43 | 1.41 | 4.65 | 0.27 | 0.88 | 0.92 |
| Stereo | 2 | 16 | 4.28 | −3.68 | 5.72 | 4.34 | 3.80 | 5.77 | 0.67 | 0.60 | 0.90 |

**Table 4.** 2D geo-location accuracy analysis using different adjustment settings for the image 1 of the Innsbruck stereo test set. Given are the mean and the standard deviation in pixels ([†] ICPs equal GCPs).

| Type | GCPs | ICPs | Mean [pxl] | | STD [pxl] | |
|---|---|---|---|---|---|---|
| | | | Across | Along | Across | Along |
| constant terms | 1 | 29 | 1.23 | 0.32 | 0.97 | 0.78 |
| | 10 | 20 | −0.14 | −0.23 | 1.04 | 0.80 |
| | 30 | 30 [†] | 0.00 | 0.00 | 0.98 | 0.77 |
| constant and linear terms | 4 | 26 | −1.90 | −1.27 | 1.21 | 1.08 |
| | 10 | 20 | 0.09 | −0.46 | 1.04 | 0.79 |
| | 30 | 30 [†] | 0.00 | 0.00 | 0.90 | 0.59 |

**Table 5.** 2D geo-location accuracy analysis for all six images of the Ljubljana test site given in pixels. For each image, the first row shows the residual errors based on GCPs and the second row the residual errors for the ICPs. After adjustment the mean STDs are 0.45 pixels for GCPs and 0.73 pixels for ICPs.

| ID | Type | Initial [pxl] | | | | Adjusted [pxl] | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **Mean** | | **STD** | | **Mean** | | **STD** | |
| | | **Across** | **Along** | **Across** | **Along** | **Across** | **Along** | **Across** | **Along** |
| 1 | GCP | 1.68 | −0.10 | 0.68 | 0.67 | 0.05 | 0.00 | 0.42 | 0.51 |
| | ICP | 1.56 | 0.02 | 0.92 | 0.61 | −0.23 | 0.09 | 0.68 | 0.72 |
| 2 | GCP | 2.85 | 2.38 | 0.85 | 0.54 | 0.08 | 0.07 | 0.40 | 0.46 |
| | ICP | 2.49 | 2.68 | 1.22 | 0.77 | −0.31 | 0.26 | 0.88 | 0.73 |
| 3 | GCP | 1.54 | −0.64 | 0.69 | 0.51 | 0.04 | −0.02 | 0.45 | 0.42 |
| | ICP | 1.20 | −0.63 | 0.87 | 0.63 | −0.31 | 0.01 | 0.63 | 0.66 |
| 4 | GCP | 1.61 | 1.60 | 0.95 | 0.82 | 0.04 | 0.04 | 0.48 | 0.67 |
| | ICP | 1.14 | 1.91 | 1.29 | 0.63 | −0.42 | 0.31 | 0.87 | 0.74 |
| 5 | GCP | −2.87 | 11.72 | 0.49 | 0.42 | −0.09 | 0.32 | 0.49 | 0.37 |
| | ICP | −3.38 | 11.56 | 0.68 | 0.76 | −0.60 | 0.18 | 0.68 | 0.68 |
| 6 | GCP | −2.82 | 12.13 | 0.47 | 0.56 | −0.09 | 0.33 | 0.47 | 0.46 |
| | ICP | −3.11 | 12.14 | 0.62 | 0.99 | −0.37 | 0.27 | 0.59 | 0.94 |

Next, the 3D geo-location accuracy using the initial and optimized sensor models is assessed, based on the protocol defined in Section 4.1. The RMS values of the 3D point residuals which were achieved for East, North, Height and their 3D length are summarized in Table 6. These values indicate the 3D mapping accuracy that is feasible when employing the stereo or triplet models for 3D reconstruction. First, it is obvious that the initial accuracy of the rational polynomial models yields 3D displacements in the order of several meters, which is clearly beyond the aspired precision. Next, the values given in the table show that the stereo intersection angle $\delta$, as an equivalent for the base-to-height ratio, has a predominant impact onto the 3D geo-location accuracy. Higher accuracies are achieved for image pairs covering larger stereo intersection angles and vice versa. This aspect is depicted in Figure 17 which shows the RMS values of the 3D length residuals plotted versus the convergence angles. The red curves give the theoretical 3D errors that would result from a given 2D measurement error in meters.
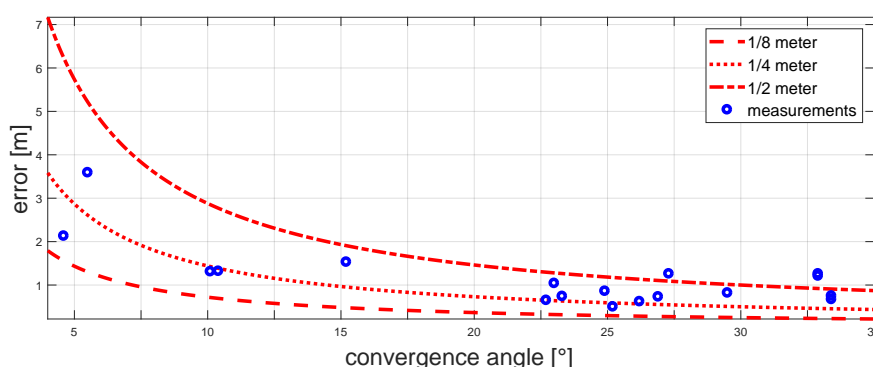


**Figure 17.** The 3D error is plotted versus the convergence angle, which are in the range from $4°$ to $35°$, assuming an 2D location measurement error of $1/8$, $1/4$ and $1/2$ m. On top of that the RMS values of the adjusted lengths (cf. Table 6) are shown in blue.

**Table 6.** 3D RMS residuals achieved from all initial as well as adjusted sensor models given in meters (with $\delta$ being the convergence angle). Values are given for East (E), North (N), Height (H) and their 3D Length (L).

| Study Area | Model | $\delta$ [°] | GCPs | RMS Initial [m] | | | | RMS Adjusted [m] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | E | N | H | L | E | N | H | L |
| Innsbruck Stereo | 1–2 | 26.9 | 30 | 4.85 | 2.07 | 3.45 | 6.28 | 0.38 | 0.23 | 0.59 | 0.73 |
| Innsbruck Triplet East | 1–2 | 10.1 | 20 | 5.26 | 4.78 | 1.46 | 7.26 | 0.33 | 0.46 | 1.18 | 1.31 |
| | 2–3 | 23.3 | 20 | 5.44 | 5.10 | 2.79 | 7.96 | 0.29 | 0.39 | 0.56 | 0.74 |
| | Triplet | 33.4 | 20 | 5.33 | 5.14 | 2.31 | 7.76 | 0.28 | 0.39 | 0.47 | 0.67 |
| Innsbruck Triplet Center | 1–2 | 24.9 | 16 | 3.48 | 1.68 | 6.09 | 7.21 | 0.36 | 0.19 | 0.76 | 0.86 |
| | 2–3 | 4.6 | 16 | 3.85 | 2.07 | 4.38 | 6.19 | 0.59 | 0.36 | 2.01 | 2.13 |
| | Triplet | 29.5 | 16 | 3.41 | 1.64 | 5.83 | 6.95 | 0.33 | 0.17 | 0.73 | 0.82 |
| Innsbruck Triplet West | 1–2 | 10.4 | 22 | 5.03 | 4.72 | 2.17 | 7.24 | 0.19 | 0.33 | 1.26 | 1.32 |
| | 2–3 | 23.0 | 22 | 5.17 | 4.88 | 3.45 | 7.90 | 0.26 | 0.27 | 0.98 | 1.04 |
| | Triplet | 33.4 | 22 | 5.09 | −4.96 | 2.97 | 7.70 | 0.21 | 0.25 | 0.67 | 0.75 |
| Trento Triplet | 1–2 | 5.5 | 21 | 4.75 | 8.14 | 6.40 | 11.39 | 0.94 | 0.88 | 3.35 | 3.59 |
| | 2–3 | 27.3 | 21 | 4.52 | 7.66 | 3.88 | 9.70 | 0.36 | 0.34 | 1.15 | 1.26 |
| | 1–3 | 32.9 | 21 | 4.60 | 7.60 | 4.16 | 9.81 | 0.42 | 0.31 | 1.15 | 1.26 |
| | Triplet | 32.9 | 21 | 4.70 | 7.66 | 4.07 | 9.87 | 0.38 | 0.31 | 1.11 | 1.21 |
| Ljubljana Stereo 1 | 1–2 | 25.2 | 28 | 0.99 | 0.71 | 2.72 | 2.98 | 0.29 | 0.30 | 0.28 | 0.50 |
| Ljubljana Stereo 2 | 3–4 | 22.7 | 30 | 1.16 | 0.58 | 2.90 | 3.17 | 0.30 | 0.32 | 0.48 | 0.65 |
| Ljubljana Stereo 3 | 5–6 | 26.2 | 30 | 1.54 | 6.58 | 1.02 | 6.83 | 0.31 | 0.37 | 0.40 | 0.62 |
| Singapore Stereo | 1–2 | 15.2 | 13 | 1.20 | 1.82 | 9.19 | 9.44 | 0.31 | 0.29 | 1.47 | 1.53 |

It is clearly visible that the two stereo pairs with small intersection angle (i.e., Innsbruck triplet center 1–3 with 4.6° and Trento triplet 1–2 with 5.5°) yield significant worse 3D geo-locations than all other stereo (and tri-stereo) constellations. For constellations involving larger stereo intersection angles the RMS values in planimetry are in a range of 0.2 m to 0.5 m. The RMS values in height are between 0.30 m and 1.5 m. The variations in height are diverse for the given test sites, which can be traced back to the accuracy of the reference data. Here, the best results are achieved for the Ljubljana test site, where highly accurate reference data is available such that the RMS in height also gets down to 0.3 m. In contrast, the Singapore test site holds rounded height reference values and thus this test site shows the highest RMS in height with 1.5 m. Overall, the sensor model adjustment results in optimized models that can be used to derive 3D information with high quality, that is, 0.3 m in planimetry and height, if the reference GCPs are also in this accuracy range.

Analog to the 2D accuracy assessment, the 3D mapping accuracy is foremost analyzed for two test sites based on GCPs and ICPs. Then, it is evaluated w.r.t. LiDAR reference data. First, the Innsbruck stereo set and second the Ljubljana multi-view stereo set are analyzed. For the Innsbruck test site, tests with respect to the utilization of a different number of GCPs and ICPs as well as with respect to the comparison of constant versus constant and linear nominator coefficients optimization are performed. The results of this analysis, including mean and standard deviation values of checkpoint residuals, are summarized in Table 7. Again, the utilization of a minimum number of GCPs yields systematic geo-location errors in East, North and height as manifested through the corresponding mean residual values. Over-determination as exemplary given, for example, by utilizing 10 GCPs, reduces these systematic errors to a more or less negligible order of magnitude and yields distinctly improved 3D RMS accuracy values, widely adequate for both optimization scenarios.

**Table 7.** 3D geo-location accuracy analysis using different adjustment settings for the Innsbruck stereo set ([†] ICPs equal GCPs). Values are given for East (E), North (N) and Height (H).

| Type | GCPs | ICPs | Mean Adjusted [m] | | | RMS Adjusted [m] | | |
|---|---|---|---|---|---|---|---|---|
| | | | E | N | H | E | N | H |
| constant terms | 1 | 29 | −0.79 | −0.07 | −0.93 | 0.92 | 0.39 | 1.13 |
| | 10 | 20 | 0.10 | −0.20 | −0.33 | 0.50 | 0.45 | 0.70 |
| | 30 | 30 [†] | 0.00 | 0.00 | 0.00 | 0.48 | 0.38 | 0.65 |
| constant and linear terms | 4 | 26 | 1.25 | −0.68 | −0.16 | 1.41 | 0.88 | 0.72 |
| | 10 | 20 | −0.02 | −0.30 | −0.21 | 0.46 | 0.48 | 0.75 |
| | 30 | 30 [†] | 0.00 | 0.00 | 0.00 | 0.40 | 0.30 | 0.61 |

The second test is based on the Ljubljana set, where the 3D geo-location accuracy is evaluated using GCPs and ICPs with the results depicted in Table 8. The initial RMS values reveal that the 2D geo-location residuals directly propagate to the resulting 3D geo-location accuracies. Thus, all stereo pairs containing the images 5 or 6 yield large displacements (also cf. Table 5). After adjustment, the majority of pairs yield high accuracies for the GCPs in planimetry of 0.2 m to 0.3 m and also in height of 0.2 m to 0.4 m. Actually, an accuracy at this level was never achieved before and is based on the highly accurate reference data. The statistics of ICPs are, as expected, a bit worse but no overfitting is observed. Figure 18 shows the 3D errors for GCPs and ICPs sorted w.r.t. the ICPs. Interestingly, there are four stereo pairs holding lower accuracy in height (and thus also in the 3D length), namely the pairs 1–3, 2–4, 3–5 and 4–6. When compared to the acquisition disposition depicted in Figure 6 all those pairs are pure across track images from two adjacent orbits with small convergence angles of about 12.5°, which is the reason for the poor height estimates .
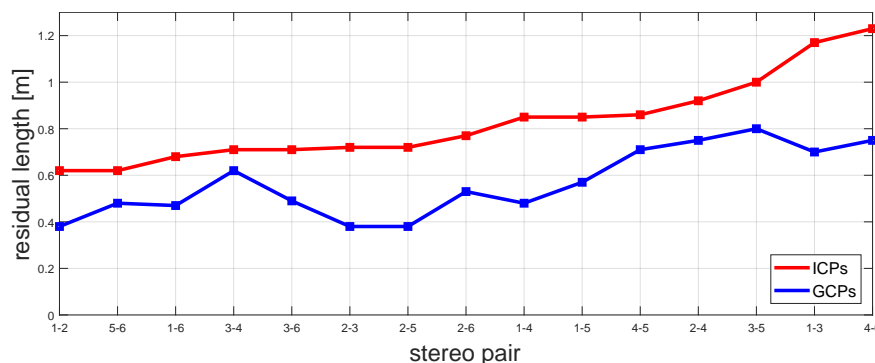


**Figure 18.** The 3D length discrepancies for the Ljubljana set sorted from lowest to highest ICP error. ICP values are given in red and corresponding GCP values in blue.

*5.2. Digital Surface Model Generation*

Before explaining the resulting DSM quality a very interesting point about epipolarity is discussed, which is neglected in other literature. As explained a high-quality epipolar rectification is a pre-requisite for 1D image matching, like the semi-global matching. However, some may think that a Pléiades stereo pair is consistent even without sensor model adjustment. Thus, it may be sufficient to extract a DSM before adjustment and later shift this DSM to some reference system (cf. Reference [5]). Therefore, a test is conducted to determine if the initial sensor models are accurate enough to get epipolar images. A number of TPs is measured in the original images and projected to the epipolar images. There, the distance orthogonal to the epipolar direction (i.e., the column direction) is gathered, which should be zero in the optimal case. Table 9 shows the statistics of the deviation of tie-points orthogonal to the related epipolar direction before and after sensor model adjustment for the Innsbruck East Triplet. While the STD values only reduce a bit, a significant change in the mean values is observed. Especially in the pair 2–3 the initial mean value is −0.66 pixel, thus more than a half-pixel off. It is known

that 1D matching degrades drastically if the correct match is not located in the current 1D search space. Overall, this is a proof that sensor models should be adjusted before further 3D processing. In cases where no GCPs are available the sensor model adjustment should be performed with TPs (cf. Reference [5]). Each TP between two images yield four additional equations in the non-linear equation system (cf. Section 4.1) but also three additional unknowns (i.e., the 3D location of that point). TPs can be automatically derived for such stereo images by, for instance, using the *scale invariant feature transform (SIFT)* [127], *speeded up robust features (SURF)* [128] or similar techniques [129]. In this case, a block adjustment can be performed for multiple images and thus for multiple RPC sensor models simultaneously.

**Table 8.** 3D RMS residuals using GCPs and ICPs for the Ljubljana test set given in meters. The adjustment is based on GCPs and evaluated on GCPs and ICPs. Values are given for East (E), North (N), Height (H) and their 3D Length (L).

| Model | Type | RMS Initial [m] | | | | RMS Adjusted [m] | | | |
|-------|------|------|------|------|------|------|------|------|------|
|       |      | E    | N    | H    | L    | E    | N    | H    | L    |
| 1–2   | GCP  | 1.01 | 0.65 | 2.61 | 2.87 | 0.18 | 0.24 | 0.22 | 0.38 |
|       | ICP  | 0.96 | 0.77 | 2.87 | 3.12 | 0.34 | 0.38 | 0.28 | 0.62 |
| 1–3   | GCP  | 0.89 | 0.34 | 1.17 | 1.51 | 0.21 | 0.25 | 0.62 | 0.70 |
|       | ICP  | 0.82 | 0.40 | 1.41 | 1.68 | 0.36 | 0.50 | 1.00 | 1.17 |
| 1–4   | GCP  | 0.96 | 0.56 | 1.59 | 1.94 | 0.21 | 0.28 | 0.33 | 0.48 |
|       | ICP  | 0.90 | 0.65 | 1.95 | 2.24 | 0.41 | 0.41 | 0.63 | 0.85 |
| 1–5   | GCP  | 0.43 | 3.80 | 3.24 | 5.01 | 0.20 | 0.23 | 0.48 | 0.57 |
|       | ICP  | 0.62 | 3.92 | 3.73 | 5.45 | 0.40 | 0.41 | 0.62 | 0.85 |
| 1–6   | GCP  | 0.68 | 3.44 | 8.78 | 9.45 | 0.20 | 0.24 | 0.36 | 0.47 |
|       | ICP  | 0.68 | 3.51 | 8.93 | 9.62 | 0.38 | 0.45 | 0.33 | 0.68 |
| 2–3   | GCP  | 1.24 | 0.62 | 3.39 | 3.66 | 0.20 | 0.23 | 0.23 | 0.38 |
|       | ICP  | 1.15 | 0.71 | 3.79 | 4.02 | 0.42 | 0.35 | 0.46 | 0.72 |
| 2–4   | GCP  | 1.43 | 0.60 | 3.46 | 3.79 | 0.19 | 0.32 | 0.65 | 0.75 |
|       | ICP  | 1.33 | 0.62 | 3.89 | 4.16 | 0.49 | 0.36 | 0.69 | 0.92 |
| 2–5   | GCP  | 0.65 | 3.60 | 4.47 | 5.78 | 0.22 | 0.21 | 0.22 | 0.38 |
|       | ICP  | 0.90 | 3.66 | 4.00 | 5.50 | 0.47 | 0.37 | 0.41 | 0.72 |
| 2–6   | GCP  | 1.12 | 2.75 | 5.51 | 6.26 | 0.20 | 0.24 | 0.43 | 0.53 |
|       | ICP  | 1.07 | 2.88 | 5.57 | 6.36 | 0.48 | 0.43 | 0.42 | 0.77 |
| 3–4   | GCP  | 1.19 | 0.55 | 2.74 | 3.04 | 0.22 | 0.29 | 0.51 | 0.62 |
|       | ICP  | 1.11 | 0.62 | 3.12 | 3.36 | 0.41 | 0.39 | 0.43 | 0.71 |
| 3–5   | GCP  | 0.85 | 4.83 | 7.02 | 8.56 | 0.24 | 0.27 | 0.71 | 0.80 |
|       | ICP  | 0.79 | 4.91 | 7.47 | 8.97 | 0.32 | 0.38 | 0.87 | 1.00 |
| 3–6   | GCP  | 2.53 | 3.77 | 12.79 | 13.57 | 0.21 | 0.22 | 0.38 | 0.49 |
|       | ICP  | 2.40 | 3.80 | 12.82 | 13.58 | 0.38 | 0.44 | 0.41 | 0.71 |
| 4–5   | GCP  | 2.76 | 2.74 | 10.43 | 11.13 | 0.28 | 0.28 | 0.59 | 0.71 |
|       | ICP  | 2.92 | 2.81 | 9.80 | 10.61 | 0.41 | 0.38 | 0.66 | 0.86 |
| 4–6   | GCP  | 3.61 | 1.49 | 14.00 | 14.54 | 0.27 | 0.31 | 0.63 | 0.75 |
|       | ICP  | 3.38 | 1.68 | 13.64 | 14.15 | 0.62 | 0.49 | 0.94 | 1.23 |
| 5–6   | GCP  | 1.48 | 6.59 | 0.93 | 6.81 | 0.28 | 0.22 | 0.32 | 0.48 |
|       | ICP  | 1.63 | 6.57 | 1.13 | 6.86 | 0.31 | 0.37 | 0.40 | 0.62 |
| all   | GCP  | 0.60 | 2.58 | 3.13 | 4.10 | 0.21 | 0.24 | 0.24 | 0.40 |
|       | ICP  | 0.62 | 2.52 | 3.35 | 4.24 | 0.41 | 0.40 | 0.31 | 0.57 |
|       | both | 0.61 | 2.55 | 3.22 | 4.16 | 0.29 | 0.30 | 0.23 | 0.48 |

**Table 9.** Statistics on the deviation of tie-points orthogonal to the epipolar direction given in pixels for the Innsbruck East Triplet.

| Model | TPs | Before Adjustment [pxl] | | After Adjustment [pxl] | |
|---|---|---|---|---|---|
| | | Mean | STD | Mean | STD |
| 1–2 | 20 | −0.29 | 0.58 | −0.04 | 0.48 |
| 2–3 | 20 | −0.66 | 1.05 | 0.03 | 0.74 |

Next, for all test sites dense DSMs were extracted using the Pléiades data and the presented workflow. The following accuracy assessment is based on the protocol defined in Section 4.2.6. The evaluation is limited to the three prime test sites Innsbruck, Trento and Ljubljana. For image matching, the parameters as given in Table 10 are used for all experiments. The two sets describe the standard version and the more accurate version. The main differences are the cost functions and the search sizes of intermediate and final pyramid levels. Due to the search sizes, the accurate version is significantly slower and thus for some application the standard version is useful as well. In the given evaluation, the settings of the accurate version are used. The matching and initial DSM generation are performed with the original resolution of 0.5 m. The downsampling to 1.0 m GSD is then performed in the fusion process by using a kernel size of 3 × 3 pixel and a step size of 2 in East and North direction. This downsampling is done to allow a direct comparison to the LiDAR data which is also given with 1.0 m GSD. For visual comparison and analysis, subsets of those test sites are presented in this section.

**Table 10.** Main parameters used for stereo matching: Two sets are defined, namely the standard version and the more accurate version. Differences in the parameters are marked in bold face.

| Parameter | Standard Version | Accurate Version |
|---|---|---|
| Matcher | SGM | SGM |
| Cost function | Census 9 × 9 pixel | **Census 9 × 9 pixel and SAD 3 × 3 pixel** |
| Prediction | SRTM DEM | SRTM DEM |
| Penalties | $P_1 = 0.4$, $P_2' = 1.5$ | $P_1 = 0.4$, $P_2' = 1.5$ |
| Pyramid levels | 4 | 4 |
| Search size | 41/15/15 | **41/31/31** |
| Cost threshold | 0.5/0.75/0.75 | 0.5/0.75/0.75 |
| Backmatching distance threshold | 1.5 | 1.5 |
| Peak filter size | 20/25/25 | 20/25/25 |

Figure 19 shows a subset of the Trento test site, which covers a hospital and its surrounding area of $360 \times 360$ m$^2$. A Pléiades ortho-image, the corresponding LiDAR DSM, the stereo-derived DSMs as well as the triplet-derived DSM are illustrated. The terrain heights of the DSMs are scaled between 240 m (black) and 300 m (white). Infrastructural changes due to ongoing construction activities since the LiDAR acquisition can be seen. For instance, there is a new helipad and a parking lot, which previously used to be a park. The stereo-derived DSMs, in general, are affected by occlusion areas, which use to increase with increasing stereo intersection angle and which almost vanish in the triplet-derived DSM. The DSM derived from the 1–2 Pléiades stereo pair looks different in comparison to the others and seems to be more similar w.r.t. the LiDAR model. This is due to the small stereo intersection angle, which implies higher image similarity and thus a higher performance of the stereo matching with reduced occlusion areas. However, due to this weak geometric disposition, the height accuracy of this DSM is worse than for the other DSMs.

Exemplary quantitative analysis was made through comparison of LiDAR to stereo and triplet-derived DSMs. Due to the large temporal gap between LiDAR and Pléiades data acquisition only selected areas were analyzed which are not affected by temporal change due to construction, vegetation growth or cloud cover. The results are summarized in Table 11. The STD of the height differences confirm the interdependence of stereo intersection angles and 3D mapping accuracy,

with worst accuracy achieved for the 1–2 stereo pair, that was discussed above. The triplet-derived DSM shows similar accuracy like the other stereo-derived DSMs. Nonetheless, it shows the best consistency in comparison to the LiDAR DSM when considering its visual appearance. It shows a clearly reduced amount of occlusion areas and more reliable and improved structures of buildings.
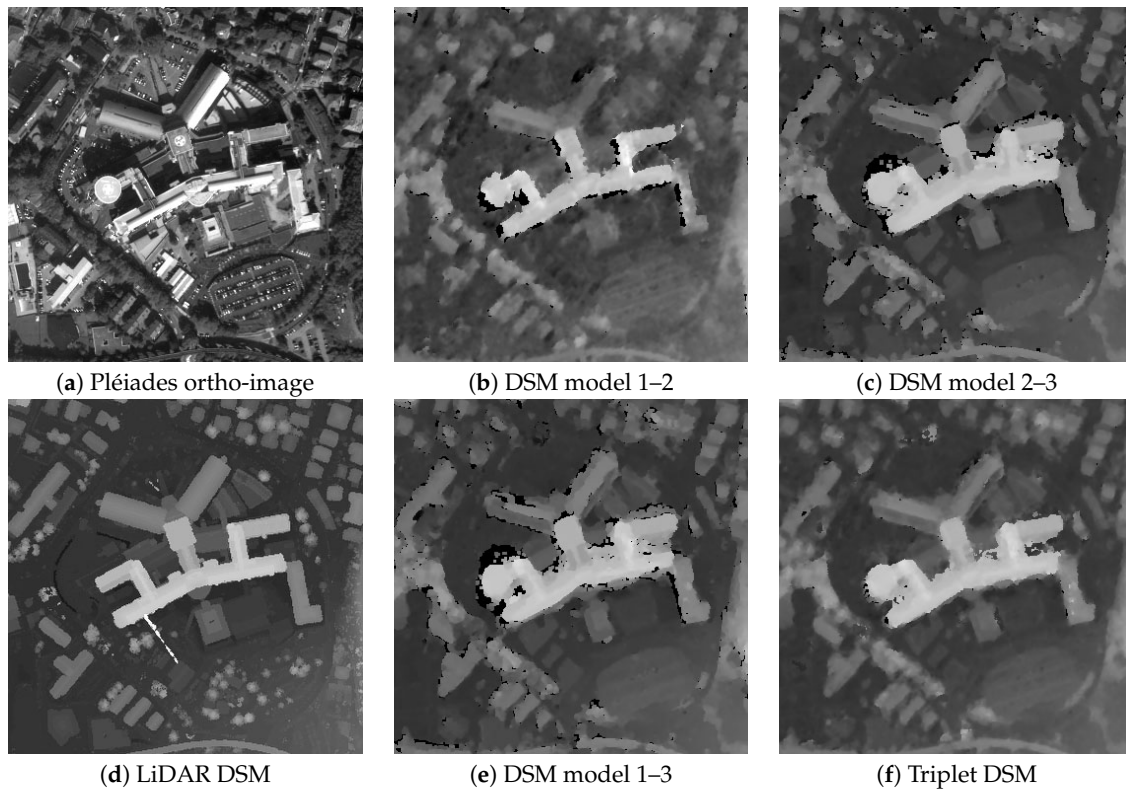


| (a) Pléiades ortho-image | (b) DSM model 1–2 | (c) DSM model 2–3 |
| --- | --- | --- |
| (d) LiDAR DSM | (e) DSM model 1–3 | (f) Triplet DSM |

**Figure 19.** Detail view of a hospital at the Trento test site.

Figure 20 shows a subset of the Innsbruck test site, which covers an urban region of $1950 \times 1000$ m$^2$. It illustrates an ortho-image, a LiDAR DSM, a stereo DSM generated by the commercial software Geomatica 2013 by PCI Geomatics (denoted as 2D matching DSM) and the stereo-derived DSM as generated by our workflow (proposed DSM). The comparison to this old PCI version is just performed to emphasize the influence of the stereo matching algorithm as in this version a 2D matcher was implemented. Newer versions also use SGM and thus show similar results than ours. The DSM height values are scaled within 620 m to 680 m. When comparing LiDAR and Pléiades DSMs, temporal changes again occur, like a new residential area in the upper-left image area. It is obvious that the proposed workflow preserves 3D breaklines better than the workflow based on 2D matching. For instance, large buildings are decently reconstructed while they are missing in the 2D matching based DSM.

**Table 11.** Statistics of height differences between LiDAR and Pléiades DSMs on non-forest regions.

| Study Area | Model | Mean [m] | STD [m] | Area [km$^2$] |
| --- | --- | --- | --- | --- |
| Trento Triplet | 1–2 | −0.62 | 4.53 | 10.8 |
|  | 2–3 | −0.55 | 1.97 | 10.8 |
|  | 1–3 | −0.53 | 2.25 | 10.8 |
|  | Triplet | −0.56 | 1.99 | 10.8 |
| Innsbruck Stereo | 1–2 | 0.25 | 1.68 | 26.5 |

(**a**) Pléiades ortho-image       (**b**) LiDAR DSM

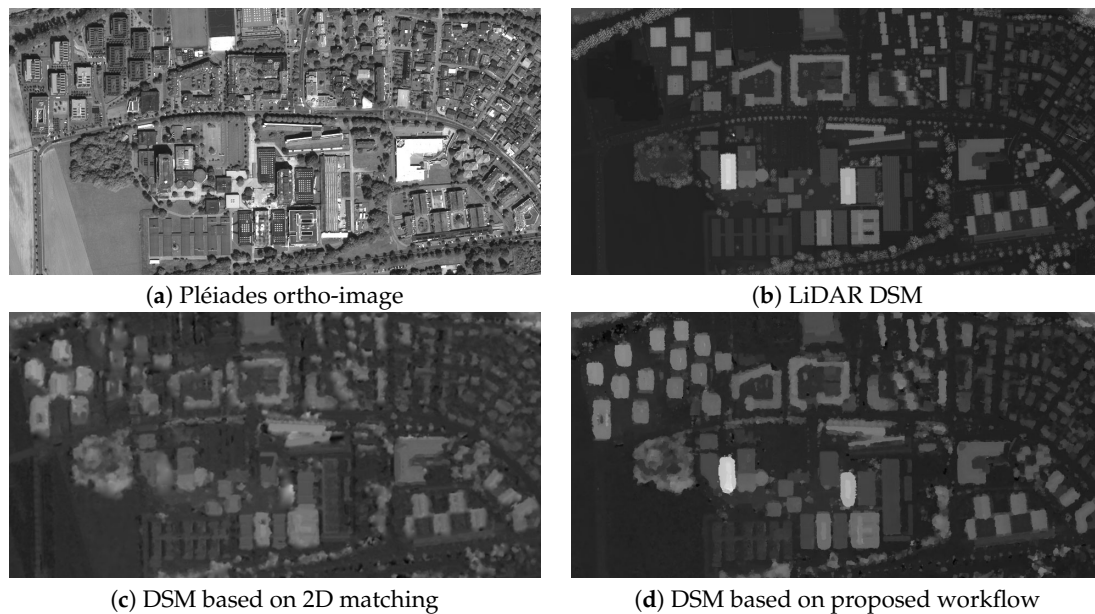(**c**) DSM based on 2D matching       (**d**) DSM based on proposed workflow

**Figure 20.** Detail view of an urban area at the Innsbruck stereo test site.

For a partly forested region at test site Innsbruck with $550 \times 550$ m$^2$ Figure 21 shows a Pléiades ortho-image as well as the height differences between the LiDAR and the Pléiades DSM, scaled from $-25$ m to $+25$ m. Thus, bright areas indicate forest clear cuts, while dark areas correspond to forests growth between the LiDAR and the Pléiades acquisition dates.



-25m      +0m      +25m

(**a**) Pléiades ortho-image       (**b**) LiDAR minus Pléiades DSM
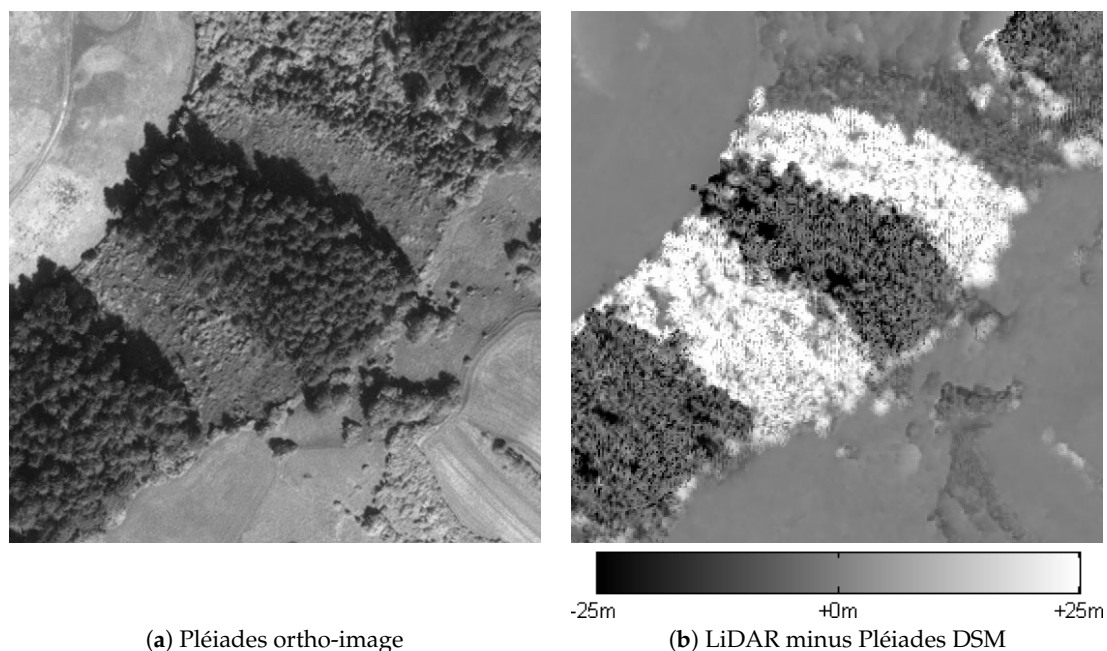
**Figure 21.** Detail view of a partly forested area at the Innsbruck stereo test site.

Analog to the Trento test site, a quantitative accuracy assessment was made for a selected area, which could be supposed to be free of temporal changes. The results are included in Table 11 and show slightly better accuracy than the Trento data sets. Overall, one very important outcome is the fact that the final DSM accuracy is worse w.r.t. the maximal possible values gathered through the sensor model optimization (cf. Table 6). Obviously, this decrease comes from the image matching step that on the one hand cannot yield the same accuracy as a manual point measurement. On the

other hand in homogenous and repetitive image regions errors will occur or incorrect matches will just be interpolated.

The last test set Ljubljana is the most interesting one, due to its multiple view geometry. The three along track stereo pairs can be used to form 15 along and across track stereo pairs as listed in Table 2. For all those 15 pairs DSMs are extracted and in addition for some selected combination of pairs. To be able to compare the coverages of different DSMs the threshold in fusion for populating an output DSM pixel is set to 1/3 of the input pixels. For example, if we have four DSMs and the fusion uses a 3 × 3 pixel kernel, then at least 12 pixels out of the 36 have to be valid such that an output is generated. This rather high threshold is chosen to allow a fair comparison of resulting nodata pixels. Note, that all remaining gaps could be filled by means of interpolation within matching or fusion. However, filling is purposely omitted to be able to recognize if multi-view images have an impact on the reconstruction completeness. Table 12 lists a multitude of statistical values, in particular minimal, maximal, mean, STD, MED, nmAD, mean absolute error (MAE), percentage of nodata values (i.e., 100% − completeness) and the convergence angles. Figure 22 complements the statistics and depicts the distributions of differences from LiDAR and Pléiades DSMs (we took the whole area of the LiDAR acquisition with 345 km$^2$ for the comparison as in References [130,131]).

**Table 12.** Accuracy analysis of the resulting DSMs in comparison to LiDAR reference data (with $\delta$ being the convergence angle). Best results are shown in bold face for single stereo and for multi stereo sets.

| Model | Min [m] | Max [m] | Mean [m] | STD [m] | MED [m] | NMAD [m] | MAE [m] | nodata [%] | $\delta$ [°] |
|---|---|---|---|---|---|---|---|---|---|
| 1–2 | −895.14 | 769.21 | −0.36 | **4.33** | 0.06 | 0.97 | 2.05 | 14.34 | 25.2 |
| 1–3 | −895.47 | 627.51 | −1.06 | 4.47 | −0.49 | 1.46 | 2.33 | **6.60** | 12.2 |
| 1–4 | −893.79 | 627.86 | −0.39 | 4.51 | 0.07 | 0.94 | **2.03** | 18.77 | 27.5 |
| 1–5 | −891.36 | 630.43 | −0.53 | 4.48 | **0.03** | 1.20 | 2.22 | 18.07 | 24.0 |
| 1–6 | −894.05 | **626.80** | −0.47 | 5.48 | −0.24 | **0.86** | 2.25 | 27.55 | 33.9 |
| 2–3 | −895.65 | 628.78 | −0.26 | 4.47 | 0.07 | 1.04 | 2.15 | 16.76 | 26.4 |
| 2–4 | −896.24 | 632.41 | −0.40 | 4.47 | 0.08 | 1.46 | 2.26 | 7.76 | 12.5 |
| 2–5 | −893.53 | 628.38 | 0.09 | 5.39 | 0.28 | 0.87 | 2.36 | 28.00 | 37.6 |
| 2–6 | −382.89 | 627.91 | −0.17 | 4.90 | **0.03** | 1.14 | 2.38 | 22.77 | 25.0 |
| 3–4 | −895.04 | 627.53 | **−0.06** | 4.49 | 0.22 | 1.01 | 2.13 | 14.82 | 22.7 |
| 3–5 | **−381.51** | 629.00 | −0.09 | 4.56 | 0.43 | 1.58 | 2.44 | 10.72 | 13.0 |
| 3–6 | −894.66 | 627.40 | −0.61 | 4.53 | −0.24 | 1.04 | 2.13 | 19.06 | 24.7 |
| 4–5 | −894.42 | 628.14 | 0.19 | 4.90 | 0.37 | 0.93 | 2.30 | 25.05 | 30.3 |
| 4–6 | −897.25 | 630.22 | −0.66 | 4.75 | −0.23 | 1.69 | 2.52 | 12.28 | 12.6 |
| 5–6 | −893.14 | 766.04 | −0.48 | 4.61 | −0.08 | 0.92 | 2.09 | 21.22 | 26.2 |
| 1–2 & 3–4 | −895.10 | 628.31 | −0.20 | 4.41 | 0.15 | 0.91 | 2.07 | 12.21 | 25.2 & 22.7 |
| 1–2 & 3–4 & 5–6 | −894.58 | 627.59 | −0.29 | 4.37 | 0.08 | 0.84 | 2.01 | 12.85 | 25.2 & 22.7 & 26.2 |
| $\delta \in [0, 20]$ | −896.42 | 626.71 | −0.51 | 4.07 | **−0.01** | 1.14 | 1.98 | 11.07 | various |
| $\delta \in [20, 30]$ | −894.17 | 627.92 | −0.37 | 4.56 | 0.05 | 0.93 | 2.16 | 7.30 | various |
| $\delta \in [30, 40]$ | −893.98 | **575.56** | **−0.05** | **2.68** | 0.16 | **0.41** | **1.00** | 55.19 | various |
| all | −894.63 | 765.43 | −0.46 | 4.74 | 0.05 | 1.09 | 2.32 | **1.12** | all |

It can be observed that there are huge outliers both positive and negative. They actually stem from the LiDAR reference data as there some height values in rivers are completely incorrect. As discussed in Section 4.2.6 those outliers lead to non-robust mean and STD estimates. The MAE measure also suffers from outliers and cannot contribute to the interpretation. Next, the plots reveal that four distributions are significantly worse (i.e., larger STD and nmAD) than all others. Those pairs are 1–3, 2–4, 3–5 and 4–6. Those pairs are exactly the four pure across track stereo pairs (cf. Figure 6) with small convergence angles. Due to the outliers, those pairs cannot be detected reliably with the STD estimate but with the nmAD. Two of these distributions show a bias in their mean value, that is, pairs 1–3 and 3–5. Again this behavior is not mapped to the mean values but to the MED values. This just emphasizes that non-robust fitting of a normal distribution based on mean and STD is not the best idea for our DSM difference models as they are non-Gaussian distributed.
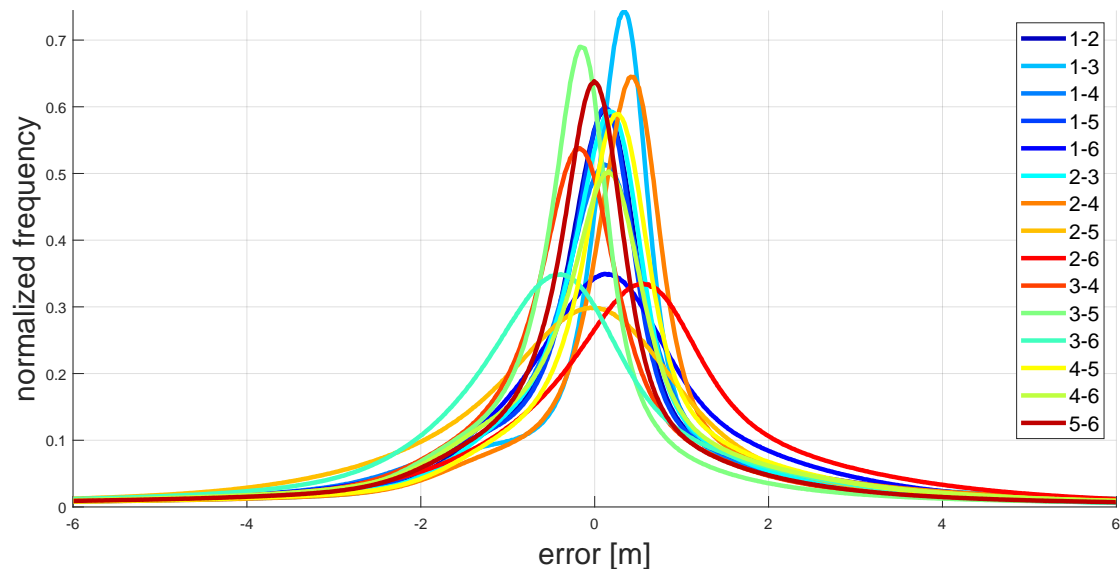
**Figure 22.** Distribution of differences from LiDAR and Pléiades DSMs for all 15 stereo pairs of the Ljubljana test site. The error range of −6 m to +6 m is shown.

As discussed above, it can be observed that small intersection angles lead to large DSM completeness (i.e., low nodata percentage) where the 1–3 pair performs best. However, despite the completeness, this pair is rather inaccurate with one of the largest nmAD values. This behavior can be traced back to the small intersection angle, such that the stereo images are more similar and thus, image matching results in a more complete disparity map. However, due to the small intersection angle, the spatial forward intersection yields a larger 3D error. Same holds, for instance, for the 2–4 pair.

We also plotted STD, nmAD and nodata percentages versus the convergence angles as depicted in Figure 23. On the left side, the plots are given using the complete error distribution including all gross outliers. On the right side, only errors with absolute values smaller than 6 m are used such that gross outliers are removed. At first, it can be seen that the STD trends change completely in these two experiments. While using all data the STD is directly correlated to the convergence angle, it is indirectly correlated when removing the gross outliers. In addition, the value range of the STD changes drastically. This behavior again shows that the STD estimate is indeed not robust by any means and should be avoided when interpreting DSM differences. The NMAD plots on the other side yield the same trend and comparable number ranges for both experiments. As expected the nmAD decreases (and thus the accuracy of the DSMs increases) with increasing convergence angles. The percentage of nodata values is directly correlated with the convergence angle and, obviously, the number of nodata values increases in the second experiments as outliers are removed.

Next, combinations of stereo pairs are processed. Groups with small intersection angles in the range of 10° to 20° (i.e., pairs 1–3, 2–4, 3–5 and 4–6), medium angles in 20° to 30° (i.e., pairs 1–2, 1–4, 1–5, 2–3, 2–6, 3–4, 3–6 and 5–6) and large angles in 30° to 40° (i.e., pairs 1–6, 2–5 and 4–5). Here, the medium angles perform best, seen on the low nmAD and low nodata values. Only smaller or only larger angles degrade the resulting DSM. When using all 15 pairs, obviously, the nodata regions are lowest with only 1.1% (i.e, a completeness of 98.9%). Note, that the reference LiDAR data holds 0.44% of nodata values, such that the real completeness is even higher than reported. Also the nmAD of all pairs is decently small, however it is even lower for, for example, the pairs in medium angle range. Here, we conclude that stereo pairs should be selected within a certain intersection angle range, where 20° to 30° seem optimal (which correspond to a B/H ratio in the range from 0.35 to 0.55). Two additional combinations are processed, namely 1–2 & 3–4 and 1–2 & 3–4 & 5–6, that is, only along track stereo pairs. As all those pairs have convergences angles around 25° the results are quite good, where the nmAD of the triple along track pair is even the best.
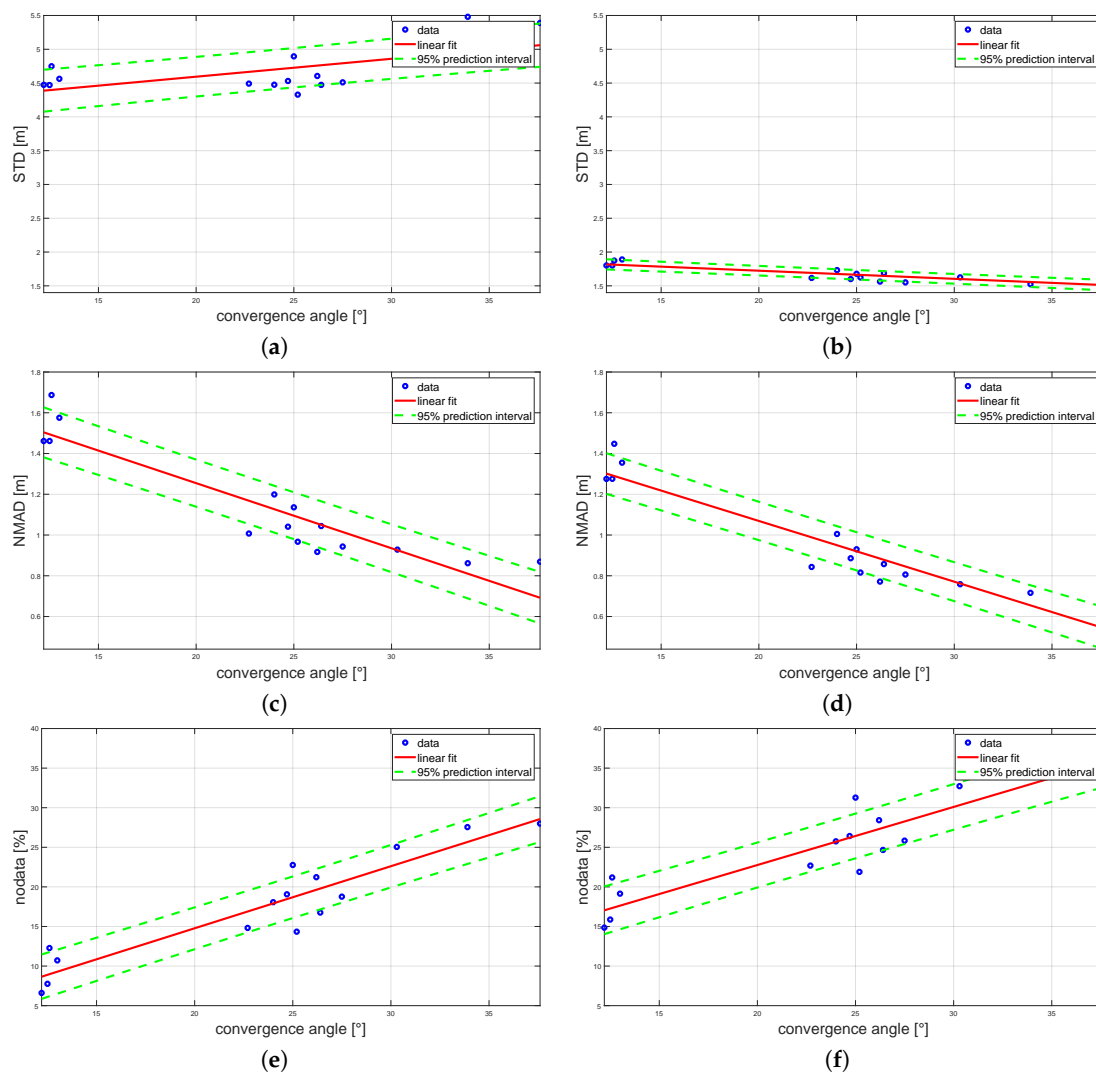
**Figure 23.** Plots showing (**a**,**b**) STD, (**c**,**d**) nmAD and (**e**,**f**) percentage of nodata values plotted versus the convergence angle for the Ljubljana test site. (**a**,**c**,**e**) are based on the complete error distribution, while (**b**,**d**,**f**) is based on a clipped distribution, where all gross outliers with an absolute value larger than 6 m are removed.

Figure 24 visualizes the differences in various DSMs. Next, to the reference LiDAR DSM the Pléiades DSM is shown based on the fusion of all stereo pairs with convergence angles in from 20° to 30°. For comparison also the DSM from the good stereo constellation 1–2 is depicted and from the worse performing constellation 1–3. At first, we see that the LiDAR model holds information like the power lines that cannot be reconstructed from Pléiades images (but nevertheless contributes to our error metric). From the Pléiades DSMs the fused one is visually a lot better, since it contains less gross errors (cf. the roof of the hall in the lower part of the image), it is more complete and it is also smoother. Comparing only the single stereo results the pair 1–2 yields a smoother surface than 1–3. Additionally, the roof structure of the central large building is not reconstructed from the 1–3 pair. Overall, this visual comparison supports our assumption that using multi-view Pléiades sets allow higher accurate 3D modeling than single stereo sets.

(**a**)  (**b**)

(**c**)  (**d**)
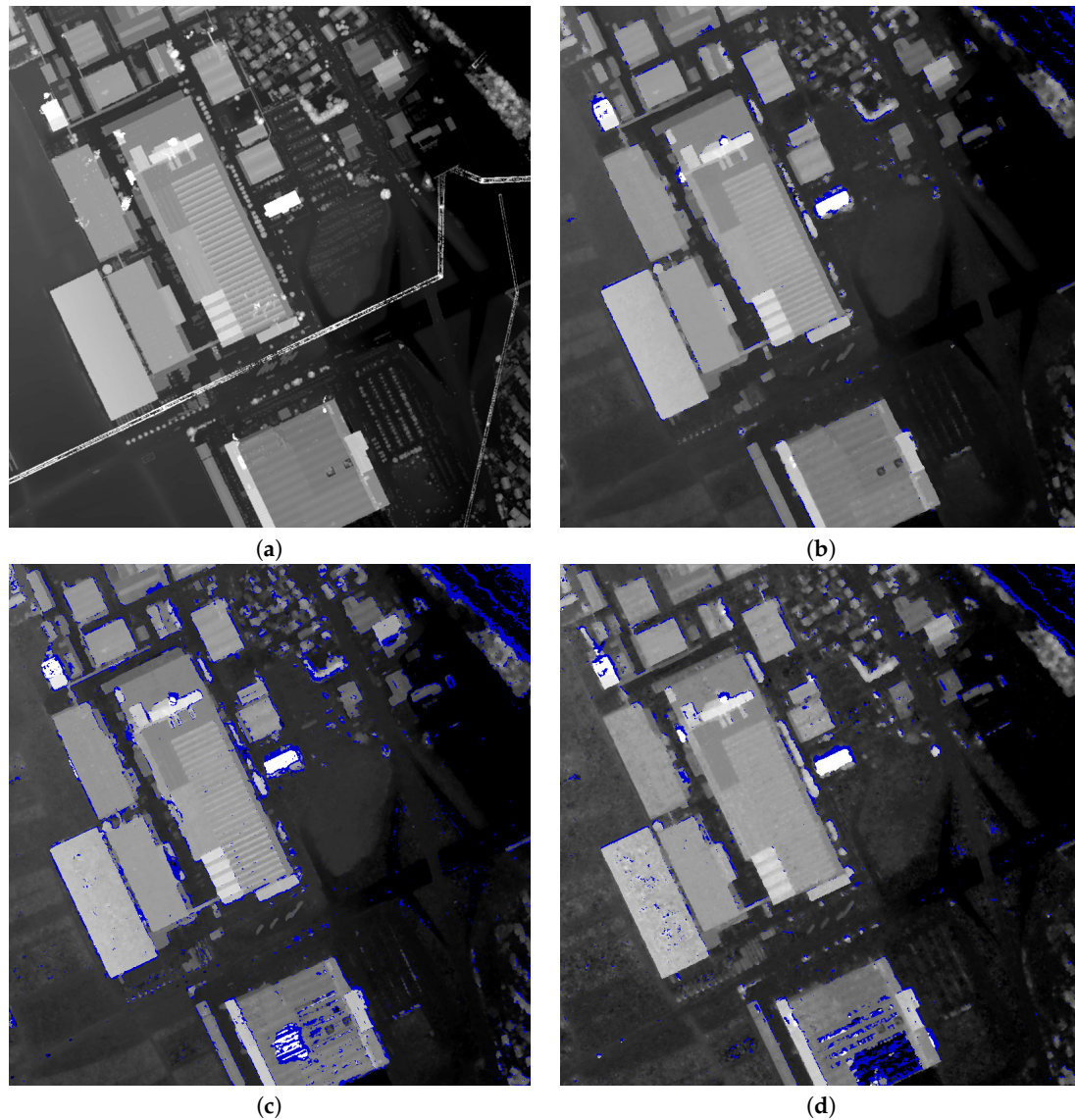
**Figure 24.** Selected DSM subsets with $800 \times 800$ m$^2$ and 1 m GSD of the Ljubljana test site. Height ranges from 410 m to 500 m. Shown are (**a**) LiDAR reference, (**b**) Pléiades DSM based on all stereo pairs with convergence angles in the range of $20°$ to $30°$, (**c**) Pléiades DSM based on stereo pair 1–2 and (**d**) Pléiades DSM based on stereo pair 1–3. Dark blue pixel indicate non-reconstructed regions.

To allow a visual interpretation of the high quality of our results, a subset of the Tian Shui test site is depicted in Figure 25 with 1 m GSD. There, the CIR ortho image is shown together with the relief shaded DSM covering $2001 \times 1701$ m$^2$ with heights from 1380 m to 1780 m. The agricultural use of the landscape is clearly visible due to the nicely reconstructed rice terraces.

### 5.3. Digital Terrain Model Generation

Tests are performed on the Innsbruck East triplet data set where the DSM is derived with 1 m GSD. For DSMs with lower or higher resolution, the processing parameters very likely have to be appropriately tuned to achieve optimal filtering results. For all tests the values given in Table 13 have been used.

**Table 13.** Parameters for terrain model extraction as used in all tests.

| Parameter | Value |
| --- | --- |
| Filter extent | 91 m |
| Height threshold | 3 m |
| Slope threshold | 30 degrees |



(**a**)  (**b**)
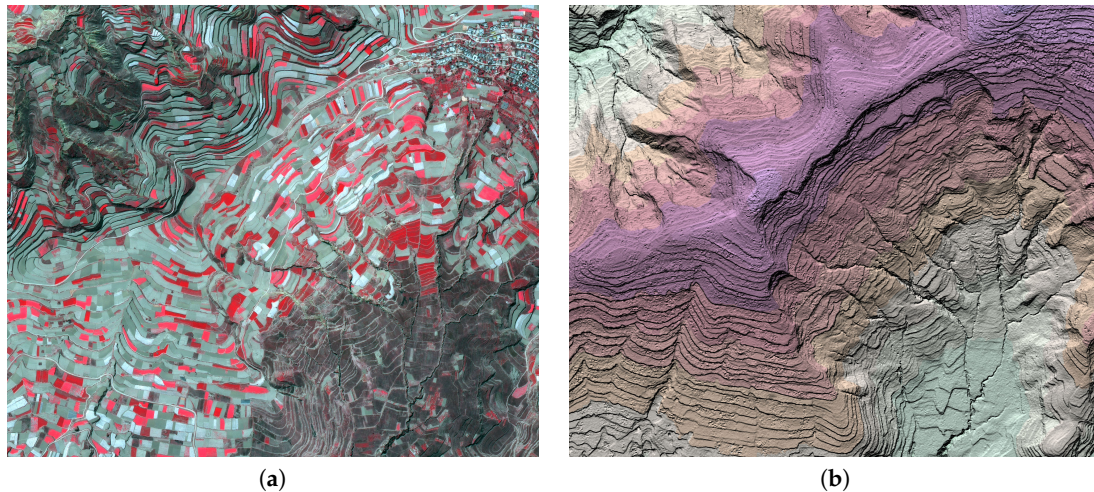
**Figure 25.** Ortho CIR image (**a**) and relief shaded DSM (**b**) are visualized for a subset of the Tian Shui test site covering 3.4 km².

### 5.3.1. Qualitative Evaluation

For visual comparison, the results of several ground pixel filtering methods are given in Figure 26 for an area of 721 × 361 m², showing a residential area at a hillside. The figure shows (a) the pansharpened true-ortho image (c) along with the input DSM. Further, it shows (b,d,e) the ground masks as resulting from DSM filtering by (b) using only two directions (i.e., left and right, which corresponds to the method in Reference [93]), (d) using 8 directions and (e) using 8 directions as well as slope dependent processing. It is clearly visible that due to the hillside location the first two methods filter too many points due to the negligence of the local terrain slope, whereas our advanced MSD method clearly yields a highly plausible and reliable segmentation. Forest areas and houses are removed while many ground points are correctly detected. Figure 27 shows a comparison of (a) 3D views of the input DSM and (b) the resulting DTM, which makes the removal of trees and houses very well visible. Figure 28 additionally shows detailed views of an urban and a suburban region.

### 5.3.2. Quantitative Evaluation

First, the Pléiades tri-stereo based DSM is compared to the LiDAR DSM (cf. Section 4.2.6). Due to the large temporal gap between LiDAR and Pléiades data acquisitions only selected areas are analyzed, which are not affected by temporal changes due to construction, vegetation growth or cloud cover. Mean values as well as STDs of the height differences are summarized in Table 14 for an area of 25.4 km². Here, the Pléiades based DSM has a bias of 0.64 m and thus, is actually too high. However, this is within the height uncertainty given for this sensor [19,26,132]. Consequently, the differences between LiDAR and Pléiades based DTMs will have the same bias, which however is not a result of the presented DTM generation algorithm.

Second, the difference of the reference LiDAR DTM to the Pléiades DTM, which has been extracted using the proposed algorithm, is analyzed. The results are given in Table 14. The mean bias between both DTMs indeed is very similar to the one achieved for the DSMs. Because not all non-ground points are perfectly removed during the Pléiades DTM generation, the achieved DTM is locally above the LiDAR DTM, resulting in an additional mean height difference of 0.11 m. However, this bias is really

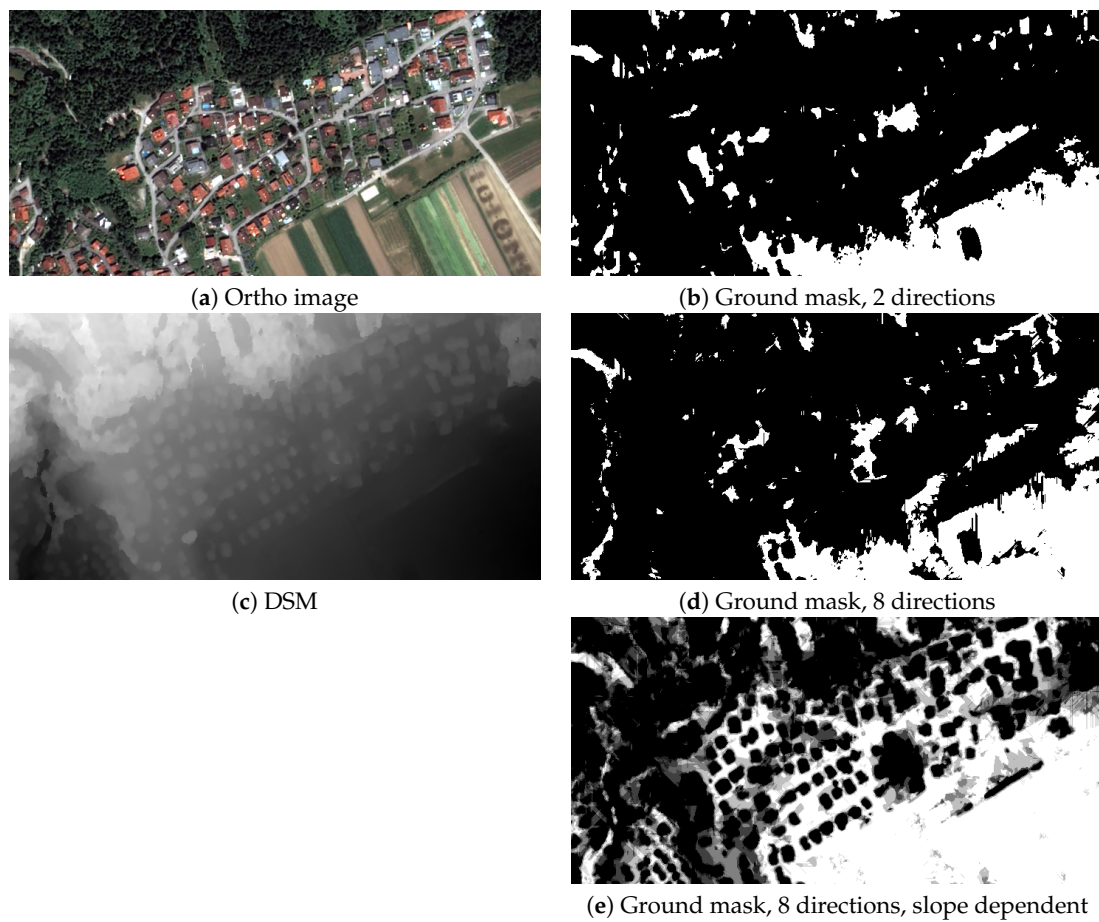small and it is below the absolute accuracy of LiDAR height measurement as well as Pléiades stereo height measurement.



(**a**) Ortho image



(**b**) Ground mask, 2 directions



(**c**) DSM



(**d**) Ground mask, 8 directions



(**e**) Ground mask, 8 directions, slope dependent

**Figure 26.** Subset of (**a**) true-ortho image, (**c**) input DSM and (**b**,**d**,**e**) together with ground masks using different methods.



(**a**)



(**b**)
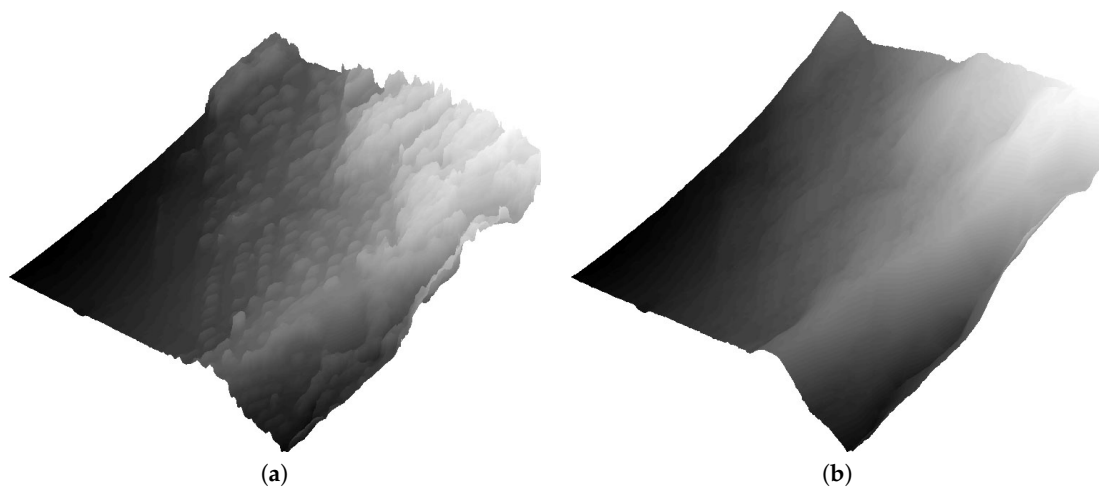
**Figure 27.** 3D view of (**a**) the DSM as shown in Figure 26 and (**b**) the resulting DTM after applying the proposed filtering method.
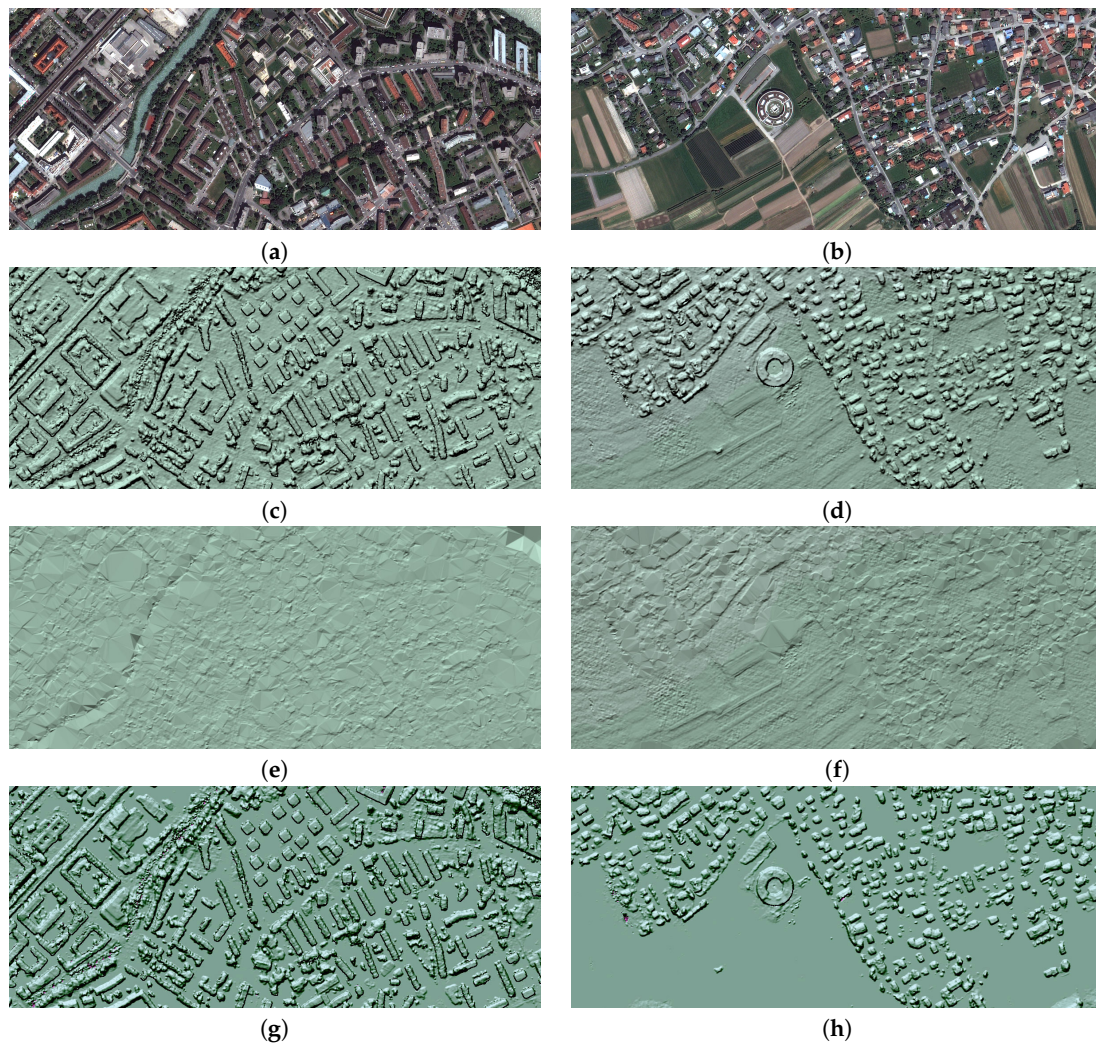
**Figure 28.** Urban subset (**left**) and suburban dataset (**right**). Shown are (**a**,**b**) ortho image, (**c**,**d**) extracted DSM, (**e**,**f**) derived DTM and (**g**,**h**) according nDSM in relief view.

**Table 14.** Statistics of height differences between LiDAR and Pléiades based DSM and DTM for the Innsbruck stereo data set.

| Model | Mean [m] | STD [m] | Area [km$^2$] |
|-------|----------|---------|-----------|
| ΔDSM | −0.64 | 2.42 | 25.4 |
| ΔDTM | −0.75 | 1.10 | 25.4 |

Figure 29 shows DSMs and DTMs as generated from LiDAR and Pléiades data for two small areas of $100 \times 100$ m$^2$ each, as well as profiles for selected objects, representing (a) a skyscraper and (b) a round building, respectively. Analysis of the plotted profiles indicates that small structures like the buildup on the first skyscraper or the hole in the center of the round building, are not reconstructed using Pléiades data. The DTM generation is able to remove buildings and in some places our DTM looks even better than the LiDAR DTM, for instance, on the left border of the round building where the LiDAR DTM is even below ground level.
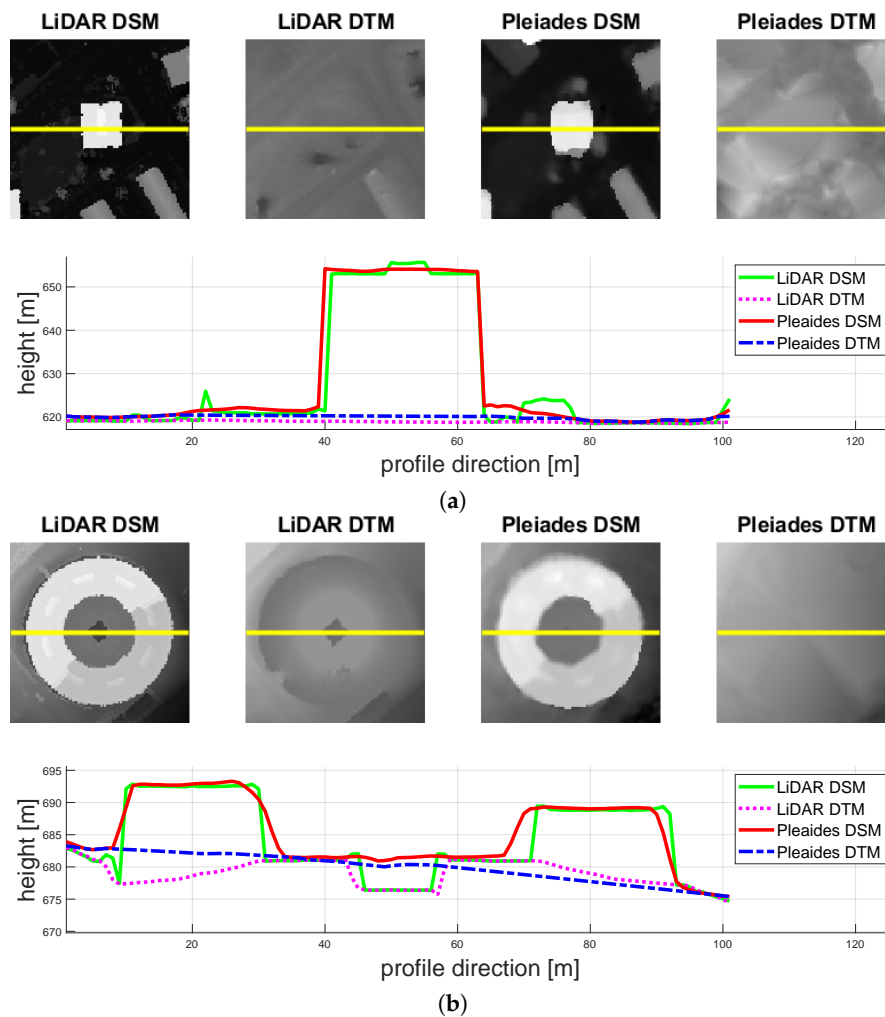
**Figure 29.** Profile comparisons of DSMs and DTMs from LiDAR and Pléiades for two small scenes, representing (**a**) a skyscraper and (**b**) a round building.

Overall, the resulting DTMs are visually appealing, which is also confirmed in the quantitative evaluation w.r.t. LiDAR reference data. Nonetheless, as long as the extraction of a DTM solely relies on the corresponding DSM, dedicated problem features and areas are inherent to the filtering algorithm being applied:

1. First, large non-ground objects may not be filtered. In particular, if an object is larger than the DTM filtering size in all eight directions, it will remain in the DTM.
2. Second, this obviously holds for larger forests, where no ground point is visible. Such forest stands will also remain in the DTM.
3. Third, the presented robust slope estimation will fail on mountain peaks such that the minimum value will be determined incorrectly. Then, the peaks will be cut off with the size of approximately half of the filter kernel size.
4. Fourth, inaccurate input DSMs directly propagate failures to the resulting DTMs. Especially local DSM outliers with low height values are classified as ground points automatically.

In the post-processing, small patches in the ground mask are removed by applying a peak filter (cf. Figure 26), since, for example, single pixels are more unreliable than others. The resulting interpolated DTM is not modified, as we want to compare the initial filtering results. A smoother surface could always be gathered by, for example, applying a Gaussian, a median or a bilateral filter [133]. An in-depth comparison and evaluation of the proposed method with several state-of-the-art methods is currently in preparation and will be soon submitted for review.

## 6. Discussion

Researchers and operators require robust and automatic workflows that yield highly accurate 3D mapping products from VHR multi-view stereo satellite data. Thus, this manuscript proposed an end-to-end workflow that yields the desired 3D mapping on the example of the Pléiades satellite constellation. The contributions w.r.t. the methodological development are as follows:

1. Introduction of the complete end-to-end 3D mapping workflow.
2. RPC-based sensor model adjustment in object-space, employing Newton's method to solve the non-linear equation systems.
3. Epipolar rectification, where the method in Reference [54] was modified and the first time applied to optical satellite images.
4. Interpolative approach in the non-linear epipolar rectification for speed-up.
5. Combined Census transform with SAD matching cost for increasing the accuracy of image matching.
6. Explicit modeling of the function which is used to vary the penalty $P_2$ in SGM.
7. Extension of classical SGM to a truncated SGM by limiting the local disparity range, resulting in significant speed-up, while yielding comparable accuracy.
8. Memory optimization in SGM by discretizing the disparity space image to 16 bits.
9. Explicit modeling of nodata values in the epipolar images during matching for speed-up.
10. Disparity prediction for increasing the robustness and efficiency in top-level image matching.
11. Subpixel normalization of the resulting disparity maps.
12. Extension of the local DSM fusion method to allow downsampling within the fusion step.
13. Novel multi-directional slope dependent DTM generation method, which extends the approach in Reference [93].

Additionally, all steps of the workflow were evaluated by in-depth accuracy assessments based on a multitude of satellite and the related metadata. The assessment itself represents a main contribution of this work. Since the presented methods were implemented within a commercially available software interested audience can directly build upon our findings. Note that the presented workflow was already successfully applied by other research groups, for instance, in References [4,110,134–136]. Especially, the DTM generation was applied in References [102,137–143].

### 6.1. Main Findings

From the methodological point of view, we stated that the presented RPC sensor model adjustment in object-space is simple and works well. In contrast to image-space based methods, no additional information like the coefficients of the affine transformation has to be stored. Furthermore, height dependency is an intrinsic part of the adjustment. Epipolar rectification is solved, the presented solution is very general and can also be used for SAR geometries. Several extensions in image matching were presented, mainly for speedup reasons. Details on the adaptive penalty paradigm were presented together with an extension to use truncated disparity space images (again for speedup). For better robustness, the concept of top-level predictions was introduced, which is based on a coarse input DSM, if existing. Subpixel refinement was discussed and a simple solution was described that yields uniform distributed subpixel fraction distribution. Storage concepts for 3D point clouds and their resampling to a regular grid were discussed as well. For DSM fusion a simple method was enhanced to also allow downsampling in the same step. Statistical measures, in particular, the MED and nmAD, were discussed and insights were given why the mean and the STD should not be used to compare DSMs. Regarding DTM generation a novel method was proposed and described in detail. In comparison to its initial publication, it was also extended. The method itself sticks out with its simplicity while yielding good results.

Also, the evaluation of all 24 Pléiades images revealed exciting insights. As expected the initial 2D and thus also 3D geo-location accuracy is inadequate for mapping. It also has been shown that even one along track stereo pair may suffer from relative geometric distortions, which was demonstrated

as deviation orthogonal to the epipolar direction. After adjustment, the residual errors dropped significantly. Here, we observed that the lower limit comes mainly from the quality of the reference GCPs and the image measurements. For the Ljubljana multi-view set, where highly precise reference information was available, a 2D and 3D accuracy was achieved as never achieved before (3D accuracies in the range of 0.2 m to 0.3 m in planimetry and 0.2 m to 0.4 m in height w.r.t. GCPs). This can be seen as empirical proof that the proposed sensor model adjustment and spatial forward intersection procedure perform as desired. During the analysis, we detected an accuracy drop within the across track stereo pairs of the Ljubljana test site, which can be traced back to the small convergence angles of those stereo pairs. Since such Pléiades multi-view constellations were never before processed they were also not reported in literature. One specific outcome of the study on the multiple view data set was that convergence angles in the range of 20° to 30°, which correspond to a B/H ratio of 0.35 to 0.55 yield best results. It seems that this convergence angle range is the optimal tradeoff in quality between image matching and spatial point intersection. Thus, in cases where many images exist, only stereo pairs with appropriate convergence angle should be used. The comparisons of LiDAR to Pléiades based DSMs showed highly accurate reconstructions, with an nmAD around 0.9 m. It also revealed that robust estimates for characterizing a Gaussian distribution have to be used in remote sensing to avoid incorrect biases when comparing DSMs. Using multiple stereo pairs for DSM generation increased the completeness and the accuracy. Terrain model extraction worked well and introduced no additional errors on the height values. Problems occur in large forests where no ground height value is visible. This aspect may degrade DTMs especially in forested hilly terrains.

Overall, the presented end-to-end workflow delivers mapping products of very high quality.

## 6.2. Future Research Goals

The most difficult blocks within the processing chain were identified as the stereo matching, the fusion process and DTM generation. The first two methods directly contribute to the quality of the resulting DSM. In stereo matching, we should try to increase the plausibility of the cost functions. A realistic way will be to learn the cost function on huge ground truth data sets based on deep learning of a convolution neural network. A second option which seems to be promising is to deeply investigate the generalized Census transform and combine it with a robust cost function that also yields useful correspondences in texture-less regions.

The fusion process should somehow globally optimize the resulting surface, while being computationally feasible. Like in SGM an RMF could be solved in a heuristic manner. Alternatively, global optimizations like in References [86,87] would be of interest as well even though they are computationally complex. Anyhow, like in photogrammetry with increasing resolutions the community will go from DSM as the desired output to a fully 3D point cloud. Thus, the fusion of multiple stereo disparities has then to be performed in 3D yielding a simplified triangulated textured point cloud representation.

As in the fusion process, DTM generation could be solved globally, for example, by a variational approach. In addition, the multi-spectral information, that is, our ortho mosaic may serve as an input in processing. A classification on the multi-spectral data would help to discriminate man-made objects and vegetation from bare earth regions.

To reduce the manual effort in the whole process automatic methods for transferring GCPs from SAR amplitude images to optical images are needed. Actually, the underlying data exists (e.g., worldwide TerraSAR-X acquisition that were used in the WorldDEM generation) and also methodologies for multi-modal image matching. In the optimal case, such a transfer workflow could be directly performed at the satellite data providers. Then, end-users would get optical images and GCPs or optical images with already optimized sensor models.

## 7. Conclusions

An end-to-end workflow for mapping with very high-resolution satellite data form the basis for any further semantic analysis. In specific, many applications in remote sensing rely on the following 3D mapping products: (1) DSM, (2) DTM, (3) nDSM and (4) ortho-rectified image mosaic. For this reason, this work described all underlying principles for satellite-based 3D mapping and proposed methods that extract all those products from multi-view stereo satellite imagery. The study was based on, but not limited to, the Pléiades satellite constellation. Besides an in-depth review of related work, the methodological part proposed solutions for each component of the end-to-end workflow. In particular, this included optimization of sensor models represented by rational polynomials, epipolar rectification, image matching, spatial point intersection, data fusion, DTM generation and ortho mosaicing. For each step implementation details were proposed and discussed. Another objective of the study was a detailed assessment of the resulting output products. Hence, a variety of test sites were chosen and data sets were gathered representing different acquisition scenarios. The assessment was based on 24 Pléiades images. First, the accuracies of the 2D and 3D geo-location were analyzed. Second, surface and terrain models were evaluated, including a critical look on the underlying error metrics. The differences between single stereo, tri-stereo and multi-view data sets were analyzed as well. Overall, 3D accuracies in the range of 0.2 m to 0.3 m in planimetry and 0.2 m to 0.4 m in height were achieved w.r.t. ground control points. Retrieved DSMs showed normalized median absolute deviations around 0.9 m in comparison to reference LiDAR data. Multi-view stereo outperformed single stereo in terms of accuracy and completeness of the resulting DSMs.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| B/H | base-to-height |
| CPU | central processing unit |
| DEM | digital elevation model |
| DSM | digital surface model |
| DTM | digital terrain model |
| GCP | ground control point |
| GPU | graphical processing unit |
| GSD | ground sampling distance |
| ICP | independent check point |
| LiDAR | light detection and ranging |
| MAE | mean absolute error |
| MED | median |
| MRF | Markov random field |
| MSD | multi-directional slope dependent |
| nDSM | normalized digital surface model |
| NMAD | normalized median absolute deviation |
| RMS | root mean square |

RPC     rational polynomial coefficients
RSG     Remote Sensing Software Graz
SAD     sum of absolute differences
SAR     synthetic aperture radar
SGM     semi-global matching
SRTM    Shuttle Radar Topography Mission
STD     standard deviation
TP      tie-point
VHR     very high resolution

## References

1.  Steinnocher, K.; Perko, R.; Hofer, M. VHR Tri-Stereo Pléiades Satellitenbilder versus Airborne Laserscan Daten—Vergleich auf Basis von abgeleiteten 3D-Gebäudemodellen. In Proceedings of the Symposium und Fachmesse für Angewandte GeoInformatik, Salzburg, Austria, 8–10 July 2014; pp. 33–38.

2.  You, Y.; Wang, S.; Ma, Y.; Chen, G.; Wang, B.; Shen, M.; Liu, W. Building detection from VHR remote sensing imagery based on the morphological building index. *Remote Sens.* **2018**, *10*, 1287. [CrossRef]

3.  Bittner, K.; D'Angelo, P.; Körner, M.; Reinartz, P. DSM-to-LoD2: Spaceborne Stereo Digital Surface Model Refinement. *Remote Sens.* **2018**, *10*, 1926. [CrossRef]

4.  Persson, H.J.; Perko, R. Assessment of Boreal Forest Height from WorldView-2 Satellite Stereo Images. *Remote Sens. Lett.* **2016**, *7*, 1150–1159. [CrossRef]

5.  Piermattei, L.; Marty, M.; Karel, W.; Ressl, C.; Hollaus, M.; Ginzler, C.; Pfeifer, N. Impact of the Acquisition Geometry of Very High-Resolution Pléiades Imagery on the Accuracy of Canopy Height Models over Forested Alpine Regions. *Remote Sens.* **2018**, *10*, 1542. [CrossRef]

6.  Piermattei, L.; Marty, M.; Ginzler, C.; Pöchtrager, M.; Karel, W.; Ressl, C.; Pfeifer, N.; Hollaus, M. Pléiades satellite images for deriving forest metrics in the Alpine region. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *80*, 240–256. [CrossRef]

7.  Bagnardi, M.; González, P.J.; Hooper, A. High-resolution digital elevation model from tri-stereo Pleiades-1 satellite imagery for lava flow volume estimates at Fogo Volcano. *Geophys. Res. Lett.* **2016**, *43*, 6267–6275. [CrossRef]

8.  Warth, G.; Braun, A.; Bödinger, C.; Hochschild, V.; Bachofer, F. DSM-based identification of changes in highly dynamic urban agglomerations. *Eur. J. Remote Sens.* **2019**, *52*, 322–334. [CrossRef]

9.  Mora, B.; Tsendbazar, N.E.; Herold, M.; Arino, O. Global Land Cover Mapping: Current Status and Future Trends. In *In Land Use and Land Cover Mapping in Europe: Practices & Trends*; Springer: Berlin, Germany, 2014; pp. 11–30.

10. Schardt, M.; Schmitt, U.; Perko, R.; Hirschmugl, M.; Ibrahim, H.; Wang, J.; Chew, P.T.; Xue, C. 3D-Pleiades Mapping for IPCC Reporting. In Proceedings of the Mapping Urban Areas from Space, Frascati, Italy, 30–31 October 2018; Volume 2.

11. Perko, R.; Hirschmugl, M.; Papst, E.; Schmitt, U. Mapping Singapore by Pléiades Stereo Data: Carbon Reporting and more. In Proceedings of the EARSeL Symposium, Bonn, Germany, 20–24 June 2016; Volume 36, p. 170.

12. Sofia, G.; Bailly, J.S.; Chehata, N.; Tarolli, P.; Levavasseur, F. Comparison of Pleiades and LiDAR digital elevation models for terraces detection in farmlands. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1567–1576. [CrossRef]

13. Rieg, L.; Klug, C.; Nicholson, L.; Sailer, R. Pléiades Tri-Stereo Data for Glacier Investigations—Examples from the European Alps and the Khumbu Himal. *Remote Sens.* **2018**, *10*, 1563. [CrossRef]

14. Belart, J.M.; Magnússon, E.; Berthier, E.; Pálsson, F.; Aðalgeirsdóttir, G.; Jóhannesson, T. The geodetic mass balance of Eyjafjallajökull ice cap for 1945–2014: Processing guidelines and relation to climate. *J. Glaciol.* **2019**, 1–15. [CrossRef]

15. Maxant, J.; Proy, C.; Fontannaz, D.; Clandillon, S.; Allenbach, B.; Yesou, H.; Battiston, S.; Uribe, C.; De Fraipont, P. Contribution of Pleiades-HR imagery for disaster damage mapping: Initial feedback over Asia, Africa, Europe or the Caribbean. In Proceedings of the EARSeL Symposium, Matera, Italy, 3–6 June 2013; Volume 33, pp. 491–498.

16. Đurić, D.; Mladenović, A.; Pešić-Georgiadis, M.; Marjanović, M.; Abolmasov, B. Using multiresolution and multitemporal satellite data for post-disaster landslide inventory in the Republic of Serbia. *Landslides* **2017**, *14*, 1467–1482. [CrossRef]

17. Bernard, M.; Decluseau, D.; Gabet, L.; Nonin, P. 3D capabilities of Pleiades satellite. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *39*, 553–557. [CrossRef]

18. Himmelreich, L.C.; Ladner, M.; Heller, A. Pléiades Tri-Stereo-Bilder im Hochgebirge—Eine Parameterstudie mit PCI Geomatics. *AGIT J.* **2017**, *3*, 153–162.

19. Perko, R.; Raggam, H.; Gutjahr, K.; Schardt, M. Assessment of the mapping potential of Pleiades stereo and triplet data. In Proceedings of the ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, Zurich, Switzerland, 5–7 September 2014; pp. 103–109.

20. Perko, R.; Raggam, H.; Gutjahr, K.; Schardt, M. Advanced DTM generation from very high resolution satellite stereo images. In Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Munich, Germany, 25–27 March 2015; Volume 2, pp. 165–172.

21. Perko, R.; Raggam, H.; Schardt, M.; Roth, P.M. Very High Resolution Mapping with the Pléiades Satellite Constellation. *Am. J. Remote Sens.* **2018**, *6*, 89–99.

22. D'Angelo, P.; Kuschk, G. Dense multi-view stereo from satellite imagery. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 6944–6947.

23. Bosch, M.; Kurtz, Z.; Hagstrom, S.; Brown, M. A multiple view stereo benchmark for satellite imagery. In Proceedings of the Applied Imagery Pattern Recognition Workshop, Washington, DC, USA, 18–20 October 2016; pp. 1–9.

24. Krauß, T.; D'Angelo, P.; Wendt, L. Cross-track satellite stereo for 3D modelling of urban areas. *Eur. J. Remote Sens.* **2019**, *52*, 89–98. [CrossRef]

25. Gong, K.; Fritsch, D. DSM Generation from High Resolution Multi-View Stereo Satellite Imagery. *Photogramm. Eng. Remote Sens.* **2019**, *85*, 379–387. [CrossRef]

26. Berthier, E.; Vincent, C.; Magnusson, E.; Gunnlaugsson, A.; Pitte, P.; Le Meur, E.; Masiokas, M.; Ruiz, L.; Pálsson, F.; Belart, J.; et al. Glacier topography and elevation changes from Pléiades very high resolution stereo images. *Cryosphere* **2014**, *8*, 4849–4883. [CrossRef]

27. Topan, H.; Jacobsen, K.; Cam, A.; Ozendi, M.; Oruç, M.; Bakioğlu, O.B.; Bayık, Ç.; Taşkanat, T. Comprehensive evaluation of Pléiades-1A bundle images for geospatial applications. *Arab. J. Geosci.* **2019**, *12*, 223. [CrossRef]

28. Jacobsen, K. Characteristics of very high resolution optical satellites for topographic mapping. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2011**, *XXXVIII-4/W19*, 137–142. [CrossRef]

29. Hirschmugl, M.; Gallaun, H.; Perko, R.; Schardt, M. Pansharpening—Methoden für digitale, sehr hoch auflösende Fernerkundungsdaten. In Proceedings of the Symposium und Fachmesse für Angewandte GeoInformatik, Salzburg, Austria, 6–8 July 2005; pp. 270–276.

30. Vivone, G.; Alparone, L.; Chanussot, J.; Dalla Mura, M.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A critical comparison among pansharpening algorithms. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2565–2586. [CrossRef]

31. Fourest, S.; Kubik, P.; Lebègue, L.; Déchoz, C.; Lacherade, S.; Blanchet, G. Star-based methods for Pleiades HR commissioning. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *39*, 513–518. [CrossRef]

32. Astrium. Pleiades Technical Documents: Pleiades User Guide V 2.0. Technical Report. 2012. Available online: https://www.intelligence-airbusds.com/en/4572-pleiades-technical-documents (accessed on 28 August 2019).

33. Poli, D.; Remondino, F.; Angiuli, E.; Agugiaro, G. Evaluation of Pléiades-1A triplet on Trento testfield. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *1*, 287–292. [CrossRef]

34. Gleyzes, A.; Perret, L.; Cazala-Houcade, E. Pleiades system fully operational in orbit. In Proceedings of the EARSeL Symposium, Matera, Italy, 3–6 June 2013; Volume 33, pp. 445–460.

35. Poli, D.; Remondino, F.; Angiuli, E.; Agugiaro, G. Radiometric and geometric evaluation of GeoEye-1, WorldView-2 and Pléiades-1A stereo images for 3D information extraction. *ISPRS J. Photogramm. Remote Sens.* **2015**, *100*, 35–47. [CrossRef]

36. Raggam, H. Surface mapping using image triplets. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 551–563. [CrossRef]

37. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2004.

38. Greenwalt, C.R.; Shultz, M.E. *Principles of Error Theory And Cartographic Applications*; Technical Report; Aeronautical Chart and Information Center: St. Louis, MO, USA, 1962.

39. Airbus Defence and Space. Pleiades Neo Leaflet. Technical Report. 2018. Available online: https://www.intelligence-airbusds.com/files/pmedia/public/r51130_9_leaflet-pleiadesneov2.pdf (accessed on 28 August 2019).

40. Dial, G.; Grodecki, J. Block adjustment with rational polynomial camera models. In Proceedings of the American Society for Photogrammetry and Remote Sensing Annual Conference, Washington, DC, USA, 22–26 April 2002; unpaginated CD ROM.

41. Grodecki, J.; Dial, G. Block adjustment of high-resolution satellite images described by rational polynomials. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 59–68. [CrossRef]

42. Jacobsen, K.; Topan, H. DEM generation with short base length Pleiades triplet. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 81–86. [CrossRef]

43. Fraser, C.S.; Hanley, H.B. Bias compensation in rational functions for IKONOS satellite imagery. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 53–57. [CrossRef]

44. Åstrand, P.J.; Bongiorni, M.; Crespi, M.; Fratarcangeli, F.; Da Costa, J.N.; Pieralice, F.; Walczynska, A. The potential of WorldView-2 for ortho-image production within the "Control with Remote Sensing Programme" of the European Commission. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *19*, 335–347. [CrossRef]

45. Aguilar, M.Á.; del Mar Saldaña, M.; Aguilar, F.J. Generation and quality assessment of stereo-extracted DSM from GeoEye-1 and WorldView-2 imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1259–1271. [CrossRef]

46. Eineder, M.; Minet, C.; Steigenberger, P.; Cong, X.; Fritz, T. Imaging geodesy—Toward centimeter-level ranging accuracy with TerraSAR-X. *IEEE Trans. Geosci. Remote Sens.* **2010**, *49*, 661–671. [CrossRef]

47. Raggam, H.; Perko, R.; Gutjahr, K.; Kiefl, N.; Koppe, W.; Hennig, S. Accuracy assessment of 3D point retrieval from TerraSAR-X data sets. In Proceedings of the European Conference on Synthetic Aperture Radar, Aachen, Germany, 7–10 June 2010; pp. 1–4.

48. Raggam, H.; Gutjahr, K.; Perko, R.; Schardt, M. Assessment of the stereo-radargrammetric mapping potential of TerraSAR-X multibeam spotlight data. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 971–977. [CrossRef]

49. Perko, R.; Koppe, W.; Raggam, H.; Gutjahr, K.; Schardt, M. Deriving accurate GCPs from TerraSAR-X stereo data over urban areas. In Proceedings of the EARSeL Symposium, Chania, Greece, 9–12 July 2018; Volume 38, p. 19.

50. Perko, R.; Raggam, H.; Gutjahr, K.; Schardt, M. Using worldwide available TerraSAR-X data to calibrate the geo-location accuracy of optical sensors. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 2551–2554.

51. Reinartz, P.; Müller, R.; Schwind, P.; Suri, S.; Bamler, R. Orthorectification of VHR optical satellite data exploiting the geometric accuracy of TerraSAR-X data. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 124–132. [CrossRef]

52. Merkle, N.; Luo, W.; Auer, S.; Müller, R.; Urtasun, R. Exploiting deep matching and SAR data for the geo-localization accuracy improvement of optical satellite images. *Remote Sens.* **2017**, *9*, 586. [CrossRef]

53. Wang, M.; Hu, F.; Li, J. Epipolar resampling of linear pushbroom satellite imagery by a new epipolarity model. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 347–355. [CrossRef]

54. Gutjahr, K.; Perko, R.; Raggam, H.; Schardt, M. The epipolarity constraint in stereo-radargrammetric DEM generation. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5014–5022. [CrossRef]

55. Fusiello, A.; Trucco, E.; Verri, A. A compact algorithm for rectification of stereo pairs. *Mach. Vis. Appl.* **2000**, *12*, 16–22. [CrossRef]

56. Fusiello, A.; Irsara, L. Quasi-euclidean uncalibrated epipolar rectification. In Proceedings of the International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008; pp. 1–4.

57. Orun, A.B. A modified bundleadjustment software for SPOT imagery and photography: Tradeoff. *Photogramm. Eng. Remote Sens.* **1994**, *60*, 1431–1437.

58. Kim, T. A study on the epipolarity of linear pushbroom images. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 961–966.

59. De Franchis, C.; Meinhardt-Llopis, E.; Michel, J.; Morel, J.M.; Facciolo, G. An automatic and modular stereo pipeline for pushbroom images. In Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Zurich, Switzerland, 5–7 September 2014; pp. 49–56.

60. De Franchis, C.; Meinhardt-Llopis, E.; Michel, J.; Morel, J.M.; Facciolo, G. Automatic sensor orientation refinement of Pléiades stereo images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 September 2014; pp. 1639–1642.

61. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [CrossRef]

62. Scharstein, D.; Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **2002**, *47*, 7–42. [CrossRef]

63. Kolmogorov, V.; Zabih, R. Computing visual correspondence with occlusions using graph cuts. In Proceedings of the IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 508–515.

64. Ranftl, R.; Pock, T.; Bischof, H. Minimizing TGV-based variational models with non-convex data terms. In Proceedings of the International Conference on Scale Space and Variational Methods in Computer Vision, Graz, Austria, 2–6 June 2013; pp. 282–293.

65. Rothermel, M.; Wenzel, K.; Fritsch, D.; Haala, N. SURE: Photogrammetric surface reconstruction from imagery. In Proceedings of the LC3D Workshop, Berlin, Germany, 4–5 December 2012; pp. 1–8.

66. Rupnik, E.; Pierrot-Deseilligny, M.; Delorme, A. 3D reconstruction from multi-view VHR-satellite images in MicMac. *ISPRS J. Photogramm. Remote Sens.* **2018**, *139*, 201–211. [CrossRef]

67. Qin, R. RPC stereo processor (RSP)—A software package for digital surface model and orthophoto generation from satellite stereo imagery. In Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Prague, Czech Republic, 12–19 July 2016; Volume 3, pp. 77–82.

68. Facciolo, G.; De Franchis, C.; Meinhardt, E. MGM: A significantly more global matching for stereovision. In Proceedings of the British Machine Vision Conference, Swansea, UK, 7–10 September 2015; pp. 1–12.

69. Drory, A.; Haubold, C.; Avidan, S.; Hamprecht, F.A. Semi-global matching: A principled derivation in terms of message passing. In Proceedings of the German Conference on Pattern Recognition, Münster, Germany, 2–5 September 2014; pp. 43–53.

70. Seki, A.; Pollefeys, M. SGM-nets: Semi-global matching with neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, 22–25 July 2017; pp. 231–240.

71. Scharstein, D.; Taniai, T.; Sinha, S.N. Semi-global stereo matching with surface orientation priors. In Proceedings of the International Conference on 3D Vision, Qingdao, China, 10–12 October 2017; pp. 215–224.

72. Zabih, R.; Woodfill, J. Non-parametric local transforms for computing visual correspondence. In Proceedings of the European Conference on Computer Vision, Stockholm, Sweden, 2–6 May 1994; pp. 151–158.

73. Hirschmüller, H.; Scharstein, D. Evaluation of Stereo Matching Costs on Images with Radiometric Differences. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 1582–1599. [CrossRef] [PubMed]

74. Shin, B.S.; Caudillo, D.; Klette, R. Evaluation of two stereo matchers on long real-world video sequences. *Pattern Recognit.* **2015**, *48*, 1113–1124. [CrossRef]

75. Klopschitz, M.; Perko, R.; Lodron, G.; Paar, G.; Mayer, H. Projected Texture Fusion. In Proceedings of the International Symposium on Image and Signal Processing and Analysis, Ljubljana, Slovenia, 18–20 September 2017; Volume 10, pp. 109–114.

76. Yoon, K.J.; Kweon, I.S. Adaptive support-weight approach for correspondence search. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *4*, 650–656. [CrossRef] [PubMed]

77. Fife, W.S.; Archibald, J.K. Improved census transforms for resource-optimized stereo vision. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 60–73. [CrossRef]

78. Ahlberg, C.; Ortiz, M.L.; Ekstrand, F.; Ekstrom, M. Unbounded Sparse Census Transform Using Genetic Algorithm. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Hilton Waikoloa Village, HI , USA, 7–11 January 2019; pp. 1616–1625.

79. Zbontar, J.; LeCun, Y. Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches. *J. Mach. Learn. Res.* **2016**, *17*, 2287–2318.

80. Mayer, N.; Ilg, E.; Hausser, P.; Fischer, P.; Cremers, D.; Dosovitskiy, A.; Brox, T. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 4040–4048.

81. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]

82. Knöbelreiter, P.; Vogel, C.; Pock, T. Self-Supervised Learning for Stereo Reconstruction on Aerial Images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4379–4382.

83. Nex, F.; Remondino, F.; Gerke, M.; Przybilla, H.J.; Bäumker, M.; Zurhorst, A. ISPRS benchmark for multi-platform photogrammetry. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *2*, 135–142. [CrossRef]

84. Schönfelder, A.; Gutjahr, K.; Perko, R.; Schardt, M. Fusion of Point Clouds derived from Aerial Images. In Proceedings of the OAGM and ARW Joint Workshop, Vienna, Austria, 10–12 May 2017; pp. 128–132.

85. Rumpler, M.; Wendel, A.; Bischof, H. Probabilistic range image integration for DSM and true-orthophoto generation. In Proceedings of the Scandinavian Conference on Image Analysis, Espoo, Finland, 17–20 June 2013; pp. 533–544.

86. Perko, R.; Zach, C. Globally Optimal Robust DSM Fusion. *Eur. J. Remote Sens.* **2016**, *49*, 489–511. [CrossRef]

87. Pock, T.; Zebedin, L.; Bischof, H. TGV-Fusion. In *Rainbow of Computer Science*; Springer: Berlin, Germany, 2011; pp. 245–258.

88. Pfeifer, N.; Mandlburger, G.; Otepka, J.; Karel, W. OPALS—A framework for Airborne Laser Scanning data analysis. *Comput. Environ. Urban Syst.* **2014**, *45*, 125–136. [CrossRef]

89. Davies, E.R. *Computer and Machine Vision: Theory, Algorithms, Practicalities*; Academic Press: Cambridge, MA, USA, 2012.

90. Sithole, G.; Vosselman, G. Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds. *ISPRS J. Photogramm. Remote Sens.* **2004**, *59*, 85–101. [CrossRef]

91. Meng, X.; Currit, N.; Zhao, K. Ground filtering algorithms for airborne LiDAR data: A review of critical issues. *Remote Sens.* **2010**, *2*, 833–860. [CrossRef]

92. Axelsson, P. DEM generation from laser scanner data using adaptive TIN models. *Int. Arch. Photogramm. Remote Sens.* **2000**, *33*, 110–117.

93. Meng, X.; Wang, L.; Silván-Cárdenas, J.L.; Currit, N. A multi-directional ground filtering algorithm for airborne LIDAR. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 117–124. [CrossRef]

94. Kraus, K.; Pfeifer, N. Advanced DTM generation from LIDAR data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2001**, *34*, 23–30.

95. Wiechert, A.; Gruber, M. DSM and true ortho generation with the UltraCam-L—A case study. In Proceedings of the American Society for Photogrammetry and Remote Sensing Annual Conference, San Diego, CA, USA, 26–30 April 2010; pp. 26–30.

96. Leberl, F.; Gruber, M.; Ponticelli, M.; Bernoegger, S.; Perko, R. The UltraCam large format aerial digital camera system. In Proceedings of the American Society For Photogrammetry & Remote Sensing, Anchorage, AK, USA, 5–9 May 2003; pp. 5–9.

97. Perko, R. Computer Vision for Large Format Digital Aerial Cameras. Ph.D. Thesis, Graz University of Technology, Graz, Austria, 2004.

98. Unger, M.; Pock, T.; Grabner, M.; Klaus, A.; Bischof, H. A variational approach to semiautomatic generation of digital terrain models. In Proceedings of the International Symposium on Visual Computing, Las Vegas, NV, USA, 30 November–2 December 2009; pp. 1119–1130.

99. Krauß, T.; Arefi, H.; Reinartz, P. Evaluation of selected methods for extracting digital terrain models from satellite born digital surface models in urban areas. In Proceedings of the International Conference on Sensors and Models in Photogrammetry and Remote Sensing, Tehran, Iran, 18–19 May 2011; Volume 5, pp. 1–7.

100. Krauß, T.; Reinartz, P. Urban object detection using a fusion approach of dense urban digital surface models and VHR optical satellite stereo data. In Proceedings of the ISPRS Istanbul Workshop, Istanbul, Turkey, 11–13 October 2010; pp. 1–6.

101. Piltz, B.; Bayer, S.; Poznanska, A. Volume based DTM generation from Very High Resolution Photogrammetric DSMs. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 83–90. [CrossRef]

102. Luethje, F.; Tiede, D.; Eisank, C. Terrain extraction in built-up areas from satellite stereo-imagery-derived surface models: A stratified object-based approach. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 9. [CrossRef]

103. Arefi, H.; D'Angelo, P.; Mayer, H.; Reinartz, P. Iterative approach for efficient digital terrain model production from CARTOSAT-1 stereo images. *J. Appl. Remote Sens.* **2011**, *5*, 1–19. [CrossRef]

104. Tian, J.; Krauss, T.; Reinartz, P. DTM generation in forest regions from satellite stereo imagery. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 401–405. [CrossRef]

105. Raggam, H.; Almer, A.; Buchroithner, M. Small to Large-Scale Mapping Using Multi-Resolution Image Data. *Int. Arch. Photogramm. Remote Sens.* **1993**, *29*, 184–191.

106. Toutin, T. State-of-the-art of geometric correction of remote sensing data: A data fusion perspective. *Int. J. Image Data Fusion* **2011**, *2*, 3–35. [CrossRef]

107. Kraus, K. *Photogrammetry: Geometry from Images and Laser Scans*; Walter de Gruyter: Berlin, Germany, 2011.

108. Campbell, J.B.; Wynne, R.H. *Introduction to Remote Sensing*; Guilford Press: New York, NY, USA, 2011.

109. Förstner, W.; Wrobel, B.P. *Photogrammetric Computer Vision*; Springer: Berlin, Germany, 2016.

110. Leopold, P.; Wang, T.; Perko, R.; Heiss, G.; Jung, M.; Oblin, A.; Zhang, Y. Comparing landslide mapping from DTM satellite derived data and field based studies of Loess sediments in Western China. *World Landslide Forum 4 Adv. Cult. Living Landslides* **2017**, *2*, 87–95.

111. Hermann, S.; Klette, R. *The Naked Truth About Cost Functions For Stereo Matching*; Technical Report Mitech-TR-33; Univ. Auckland: Auckland, New Zealand, 2009.

112. Bleyer, M.; Chambon, S. Does color really help in dense stereo matching. In Proceedings of the International Symposium 3D Data Processing, Visualization and Transmission, Paris, France, 17–20 May 2010; pp. 1–8.

113. Zhu, K.; D'Angelo, P.; Butenuth, M. Evaluation of Stereo Matching Costs on Close Range, Aerial and Satellite Images. In Proceedings of the International Conference on Pattern Recognition Applications and Methods, Vilamoura, Algarve, Portugal, 6–8 February 2012; pp. 379–385.

114. Paar, G.; Pölzleitner, W. Robust Disparity Estimation in Terrain Modeling for Spacecraft Navigation. In Proceedings of the IAPR International Conference on Pattern Recognition, The Hague, The Netherlands, 30 August–3 September 1992; Volume 11; pp. 738–741.

115. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The shuttle radar topography mission. *Rev. Geophys.* **2007**, *45*. [CrossRef]

116. 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. *ASTER Global Digital Elevation Model Version 2-Summary of Validation Results*; Technical Report; NASA: Washington, DC, USA, 2011.

117. Riegler, G.; Hennig, S.; Weber, M. WorldDEM—A novel global foundation layer. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, 183–187. [CrossRef]

118. Haller, I.; Nedevschi, S. Design of Interpolation Functions for Subpixel-Accuracy Stereo-Vision Systems. *IEEE Trans. Image Process.* **2012**, *21*, 889–898. [CrossRef]

119. Isenburg, M. LASzip: Lossless compression of LiDAR data. *Photogramm. Eng. Remote Sens.* **2013**, *79*, 209–217. [CrossRef]

120. Rousseeuw, P.J.; Croux, C. Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* **1993**, *88*, 1273–1283. [CrossRef]

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

122. Leys, C.; Ley, C.; Klein, O.; Bernard, P.; Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **2013**, *49*, 764–766. [CrossRef]

123. Fox, J.; Weisberg, S. *An R Companion to Applied Regression*; Sage Publications: Thousand Oaks, CA, USA, 2018.

124. Stein, M.L. *Interpolation of Spatial Data: Some Theory for Kriging*; Springer Science & Business Media: Berlin, Germany, 2012.

125. Weickert, J.; Welk, M. Tensor Field Interpolation with PDEs. In *Visualization and Processing of Tensor Fields*; Springer: Berlin, Germany, 2006; pp. 315–325.

126. Bredies, K.; Kunisch, K.; Pock, T. Total generalized variation. *SIAM J. Imaging Sci.* **2010**, *3*, 492–526. [CrossRef]

127. Lowe, G. SIFT—The Scale Invariant Feature Transform. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]

128. Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded up robust features. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 404–417.

129. Tuytelaars, T.; Mikolajczyk, K. Local invariant feature detectors: A survey. *Found. Trends® Comput. Graph. Vis.* **2008**, *3*, 177–280. [CrossRef]

130. Perko, R.; Hirschmugl, M.; Deutscher, J.; Schardt, M.; Hollaus, M.; Roth, P.M. Using multiple along and across track Pléiades stereo images for improved digital surface model generation. In Proceedings of the EARSeL Symposium, Salzburg, Austria, 1–4 July 2019; Volume 39, pp. 134–135.

131. Perko, R.; Schardt, M.; Piermattei, L.; Auer, S.; Roth, P.M. Multiple View Geometry in Remote Sensing: An Empirical Study based on Pléiades Satellite Images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3629–3632.

132. Stumpf, A.; Malet, J.P.; Allemand, P.; Ulrich, P. Surface reconstruction and landslide displacement measurements with Pléiades satellite images. *ISPRS J. Photogramm. Remote Sens.* **2014**, *95*, 1–12. [CrossRef]

133. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the International Conference on Computer Vision, Bombay, India, 4–7 January 1998; pp. 839–846.

134. Deutscher, J.; Perko, R.; Gutjahr, K.; Hirschmugl, M.; Schardt, M. Mapping tropical rainforest canopy disturbances in 3D by COSMO-SkyMed Spotlight InSAR-stereo data to detect areas of forest degradation. *Remote Sens.* **2013**, *5*, 648–663. [CrossRef]

135. Persson, H. Estimation of boreal forest attributes from very high resolution Pléiades data. *Remote Sens.* **2016**, *8*, 736. [CrossRef]

136. Himmelreich, L. DHM Ableitungen aus Pléiades Tri-Stereo Satellitenbildern im Hochgebirge. Digitale Höhenmodelle Verschiedener Softwareprodukte im Vergleich zu ALS Daten. Master's Thesis, University of Innsbruck, Innsbruck, Austria, 2017.

137. Mousa, Y.A.K.; Helmholz, P.; Belton, D. New DTM extraction approach from airborne images derived DSM. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42*, 75–82. [CrossRef]

138. Auer, S.; Schmitt, M.; Reinartz, P. Automatic alignment of high resolution optical and SAR images for urban areas. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, USA, 23–28 July 2017; pp. 5466–5469.

139. Misra, P.; Takeuchi, W. Digital Surface Model (DSM) Datasets for Built Structure Height Estimation over Indian Cities. In Proceedings of the International Remote Sensing Symposium, Nagoya, Japan, 17–19 May 2017; pp. 1–4.

140. Auer, S.; Reinartz, P.; Schmitt, M. Object-related alignment of heterogeneous image data in remote sensing. In Proceedings of the International Conference on Information Fusion, Cambridge, UK, 10–13 July 2018; pp. 1607–1614.

141. Panagiotakis, E.; Chrysoulakis, N.; Charalampopoulou, V.; Poursanidis, D. Validation of Pleiades Tri-Stereo DSM in Urban Areas. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 118. [CrossRef]

142. Mousa, Y.A.; Helmholz, P.; Belton, D.; Bulatov, D. Building detection and regularisation using DSM and imagery information. *Photogramm. Rec.* **2019**, *34*, 85–107. [CrossRef]

143. Bachofer, F.; Braun, A.; Adamietz, F.; Murray, S.; D'Angelo, P.; Kyazze, E.; Mumuhire, A.P.; Bower, J. Building Stock and Building Typology of Kigali, Rwanda. *Data* **2019**, *4*, 105. [CrossRef]