Roland Perko, Dipl.-Ing. Dr.techn.

Mapping with Pléiades

Master's Thesis

to achieve the university degree of

Master of Science

Master's degree programme: Computer Science

submitted to

Graz University of Technology

Supervisor

Dipl.-Ing. Dr.techn. Peter M. Roth

Institute of Computer Graphics and Vision
Head: Univ.-Prof. Dipl-Ing. Dr.techn. Dieter Schmalstieg

Graz, June 2019

This document is set in Palatino, compiled with pdfLaTeX2e and Biber.

The LaTeX template from Karl Voit is based on KOMA script and can be found online: https://github.com/novoid/LaTeX-KOMA-template

TO ILVY

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date	Signature

Abstract

An end-to-end workflow for mapping with very high resolution satellite data is the pre-requisite for any further semantic analysis. In specific, many applications in remote sensing need the following 3D mapping products: (1) digital surface model, (2) digital terrain model, (3) normalized digital surface model, and (4) ortho-rectified image mosaic. This thesis describes all underlying principles for satellite-based 3D mapping and proposes methods that extract all those products from multi-view stereo satellite imagery in the ground sampling distance of the input data. The study is based on, but not limited to, the Pléiades satellite constellation. Beside an in-depth review of related work, the methodological part proposes solutions for each component of the end-to-end workflow. In particular, this includes 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 step implementation details are proposed and discussed. Another aim of this thesis 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 meters in planimetry and 0.2 to 0.4 meters in height are achieved w.r.t. ground control points. Retrieved surface models show normalized median absolute deviations around 0.9 meters in comparison to reference LiDAR data. Multi-view stereo outperforms single stereo in terms of accuracy and completeness of the resulting surface models.

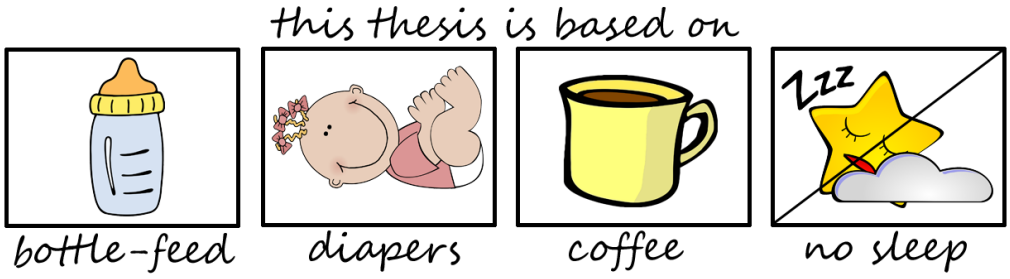
Zusammenfassung

Ein durchgängiger Workflow zur Kartierung mit Satellitendaten mit sehr hoher Auflösung stellt die Voraussetzung für jede weitere semantische Analyse dar. Viele Anwendungen in der Fernerkundung erfordern folgende 3D-Kartenprodukte: (1) digitales Oberflächenmodell, (2) digitales Geländemodell, (3) normalisiertes digitales Oberflächenmodell und (4) orthorektifiziertes Bildmosaik. Diese Arbeit beschreibt alle zugrundeliegenden Prinzipien für die satellitenbasierte 3D-Kartierung und schlägt Methoden vor, die die gewünschten Produkte aus (mehrfach) Stereo-Satellitenbildern erzeugen. Diese Abhandlung basiert auf der Pléiades-Satellitenkonstellation, ist jedoch nicht darauf beschränkt. Neben einer eingehenden Literaturstudie werden im methodischen Teil Lösungen für jede Komponente des Workflows vorgeschlagen. Dies umfasst u.a. die Optimierung von Sensormodellen, epipolare Rektifizierung, Bild-Matching, Vorwärtsschnitt, Datenfusion, Geländemodellableitung und Orthomosaikierung. Für jeden Schritt werden Implementierungsdetails vorgeschlagen und diskutiert. Ein weiteres Ziel der Arbeit ist eine detaillierte Bewertung der Ergebnisse. Hierfür werden 24 Pléiades Bilder, welche unterschiedliche Aufnahmeszenarien beschreiben, verwendet. Zunächst werden die Genauigkeiten der 2D- und 3D-Geolokalisierung analysiert. Dann werden Oberflächen- und Geländemodelle evaluiert, einschließlich eines kritischen Blicks auf die zugrundeliegenden Fehlermetriken und der Erörterung der Unterschiede zwischen einfachen und mehrfachen Stereo-Datensätzen. Im Vergleich zu Passpunkten werden 3D-Genauigkeiten von 0.2 bis 0.3 Meter in der Planimetrie und 0.2 bis 0.4 Meter in der Höhe erreicht. Oberflächenmodelle zeigen im Vergleich zu LiDAR-Referenzdaten normalisierte mittlere absolute Abweichungen um 0.9 Meter. Die Verwendung mehrerer Stereopaare erhöht die Genauigkeit und Vollständigkeit der Oberflächenmodelle.

Acknowledgments

First of all, I would like to thank my advisors Peter M. Roth, Hannes Raggam, and Mathias Schardt for supervision. Furthermore, friends and colleagues from JOANNEUM RESEARCH, TU Vienna, and German Aerospace Centre (DLR) helped by various means. Namely, I want to thank Karlheinz Gutjahr, Erich Riegler, Ursula Schmitt, Manuela Hirschmugl, Janik Deutscher, Philipp Hafner, Livia Piermattei, Markus Hollaus, Norbert Pfeifer, and Stefan Auer. Thanks also go to Martina Uray for proof reading. Concerning technical support and for his worldly wisdom I thank Gregor Stadlober.
The possibility and motivation to start and finish the study of computer science was connected with the birth of my daughter Ilvy, as I was able to apply for one year of parental and educational leave. Here, I would like to thank my bosses Mathias Schardt and Heinz Mayer for supporting this opportunity.
Finally and foremost, many thanks got to my whole family and my friends for support and in specific to my little family, i.e., my wild love Jutta and my beloved daughter Ilvy. You supported this study and without that it would not have been possible. Thank you! 🥰

Freedom! Roli
June 2019



Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

B/H	base-to-height
CPU	central processing unit
CUDA	compute unified device architecture
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

1 Introduction

For many applications highly accurate and up-to-date mapping information gathered from very high resolution (VHR) satellite stereo images is needed. Exemplary applications in the domain of remote sensing are, for instance, city modeling (Duan and Lafarge, 2016; Steinnocher, Perko, and Hofer, 2014; You et al., 2018; Bittner et al., 2018), forest assessment and biomass estimation (Persson, Wallerman, et al., 2013; Persson and Perko, 2016; Persson, 2016; Piermattei, Marty, Karel, et al., 2018; Kothencz et al., 2018; Stylianidis et al., 2019; Piermattei, Marty, Ginzler, et al., 2019), change detection (Abduelmola, 2016; Bagnardi, González, and Hooper, 2016; Warth et al., 2019), land cover and land use classification (Mora et al., 2014; Belgiu, Drăguţ, and Strobl, 2014; Schardt et al., 2018), carbon reporting (Perko, Hirschmugl, Papst, et al., 2016), farm land monitoring (Sofia et al., 2016), glacier observation (Rieg et al., 2018; Belart et al., 2019), disaster damage mapping (Maxant et al., 2013; Đurić et al., 2017), landslide mapping (Leopold et al., 2017), and mapping in general (Capaldo et al., 2012; Bernard et al., 2012; Bosch, Leichtman, et al., 2017; Himmelreich, Ladner, and Heller, 2017; Ladner, Heller, and Grillmayer, 2017; Vanderhoof and Burt, 2018). To allow semantic analysis those applications need mapping products in form of digital surface models (DSM), digital terrain models (DTM), their difference, i.e., normalized digital surface models (nDSM), and the according multi-spectral ortho-rectified image mosaics. All this information can be extracted from stereo or multi-view satellite imagery. Therefore, this thesis 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 (Perko, Raggam, Gutjahr, et al., 2014; Gutjahr et al., 2014; Perko, Raggam, Gutjahr, et al., 2015; Perko, Raggam, Schardt, et al., 2018; Perko, Hirschmugl, Deutscher, et al., 2019; Perko, Schardt, et al., 2019) which are extended, refined, and completed.

The central contributions of this thesis 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, e.g., (D'Angelo and Kuschk, 2012; Bosch, Kurtz, et al., 2016; Facciolo, De Franchis, and Meinhardt-Llopis, 2017; Qin, 2017; Rupnik, Pierrot-Deseilligny, and Delorme, 2018; Krauß, D'Angelo, and Wendt, 2018; Gong and Fritsch, 2019).
- 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 is able to 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 (Stumpf et al., 2014; Perko, Raggam, Gutjahr, et al., 2014; Berthier

et al., 2014; Qayyum, Malik, and Saad, 2015; Perko, Raggam, Schardt, et al., 2018; Panagiotakis et al., 2018; Topan et al., 2019). Note, that all presented and proposed methodologies in this thesis 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.

This thesis touches multiple core scientific disciplines, namely remote sensing, photogrammetry, and computer vision, all of them embedded into applied mathematics (cf. Figure 1.1).

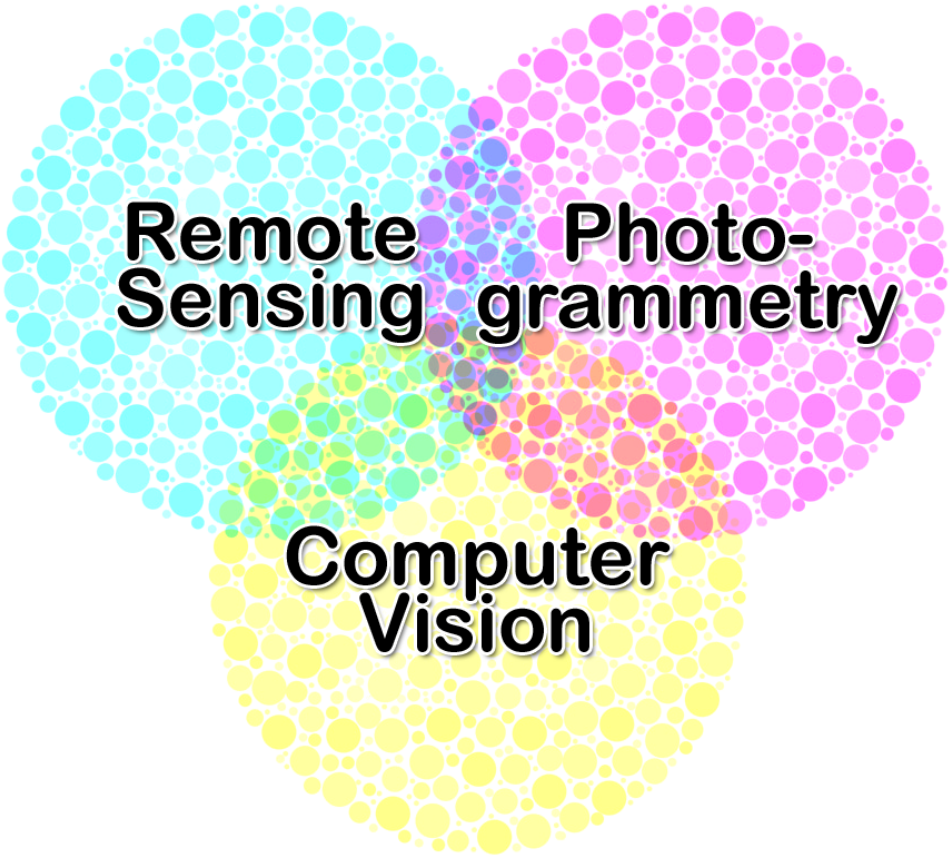
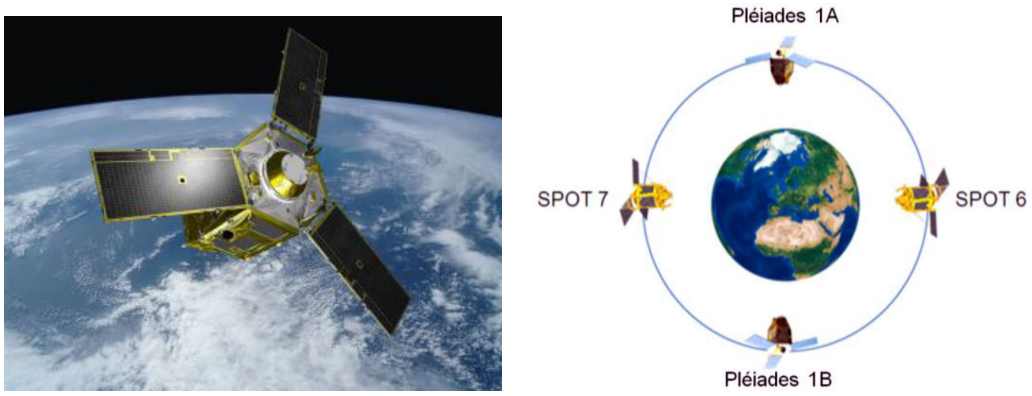


Figure 1.1: The multi-disciplinary core research topics of this thesis: remote sensing, photogrammetry, and computer vision (all embedded into applied mathematics).

1.1 The Pléiades Satellite System

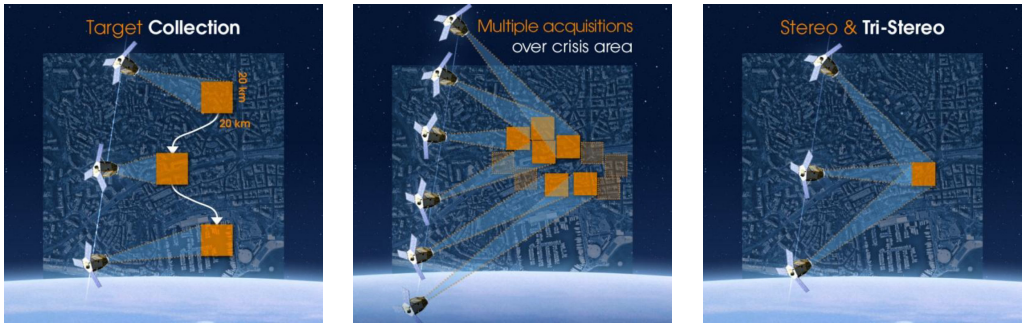
The European Pléiades satellite system is a dual system comprising the two identical satellites Pléiades-1A and Pléiades-1B. They have been launched in December 2011 and in December 2012, respectively, both providing VHR image data (where VHR is defined for resolutions or ground sampling distances (GSD) below 1 m (Jacobsen, 2011)). 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 (cf. Figure 1.2). They are supplied with a 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 (Hirschmugl et al., 2005; Vivone et al., 2015) 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 40000 pixels in across track direction. Interestingly, the Pléiades open star cluster was the best place to steer the eponymous satellite in the calibration phase (Fourest et al., 2012). Three acquisition modes are shown in Figure 1.3 and for additional information on the Pléiades sensors we refer to (Astrium, 2012; Poli et al., 2013; Gleyzes, Perret, and Cazala-Houcade, 2013; Poli et al., 2015; Topan et al., 2019).

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 over flight (single pass) through 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



(a) (b)

Figure 1.2: The Pléiades constellation: (a) artist view of the Pléiades satellite, and (b) orbital position of the satellites. Both figures are taken from (Astrium, 2012).



(a) (b) (c)

Figure 1.3: Exemplary Pléiades single pass acquisition modes: (a) target collection, (b) multiple acquisitions, and (c) stereo and tri-stereo. All figures are taken from (Astrium, 2012).

of such arrangements was made, e.g., in (Raggam, 2006), 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 (Hartley and Zisserman, 2004)) as well known from computer vision and photogrammerty. 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[1] (Astrium, 2012). Therefore, for accurate mapping the geo-location accuracy has to be refined, i.e., optimized using GCPs. Additionally, the launch of the Pléiades-Neo satellite constellation is planned between 2020 and 2022 (Airbus Defence and Space, 2018). 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 thesis presents the complete end-to-end workflow that is necessary to allow 3D mapping. The main objective is to get from a set of input images and according meta data to the mapping products in an automatic and robust manner. As sketched in Figure 1.4 four core tasks are needed to accomplish this goal. Those are sensor model adjustment, DSM generation, DTM generation and ortho-rectification, summarized as follows:

1. Sensor model geo-location accuracy, i.e., 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. Generation of digital surface models, exploiting the characteristics of the underlying sensor system, like availability of image triplets in case of Pléiades, and 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.

[1] According to (Greenwalt and Shultz, 1962) CE90 is defined as the radial error which 90% of all errors in a circular distribution will not exceed, i.e., the 90^{th} percentile of circular error distribution.

3. Extraction of a DTM as well as an nDSM from the pure DSM. This will filter man-made structures and vegetation and allows to retrieve vegetation or building heights. 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 in order to retrieve reliable terrain from such input. Algorithms as well as results which have been achieved with respect to urban mapping are discussed and validated.
4. Generation of an ortho-rectified multi-spectral image mosaic based on accurate sensor models and derived DSMs.

1.3 Software Environment

Most of the presented methods were first prototyped in Matlab. Then, they were implemented within the commercial software package *Remote Sensing Software Graz*[2] *(RSG)* by JOANNEUM RESEARCH (cf. Figure 1.5). This package is developed under C++, compute unified device architecture (CUDA), Fortran, Python, and a process control system called *PROX*. The modules proposed in this thesis are embedded within the *Space Suite*[3].

1.4 Key Publications

To show the scientific impact of this work the publications developed during the research on this thesis are listed in chronological order.

[2] http://www.remotesensing.at/en/remote-sensing-software.html

[3] https://www.remotesensing.at/fileadmin/user_upload/imported/DIGITAL/Downloads/Remote_Sensing_Software_Graz_2018.pdf

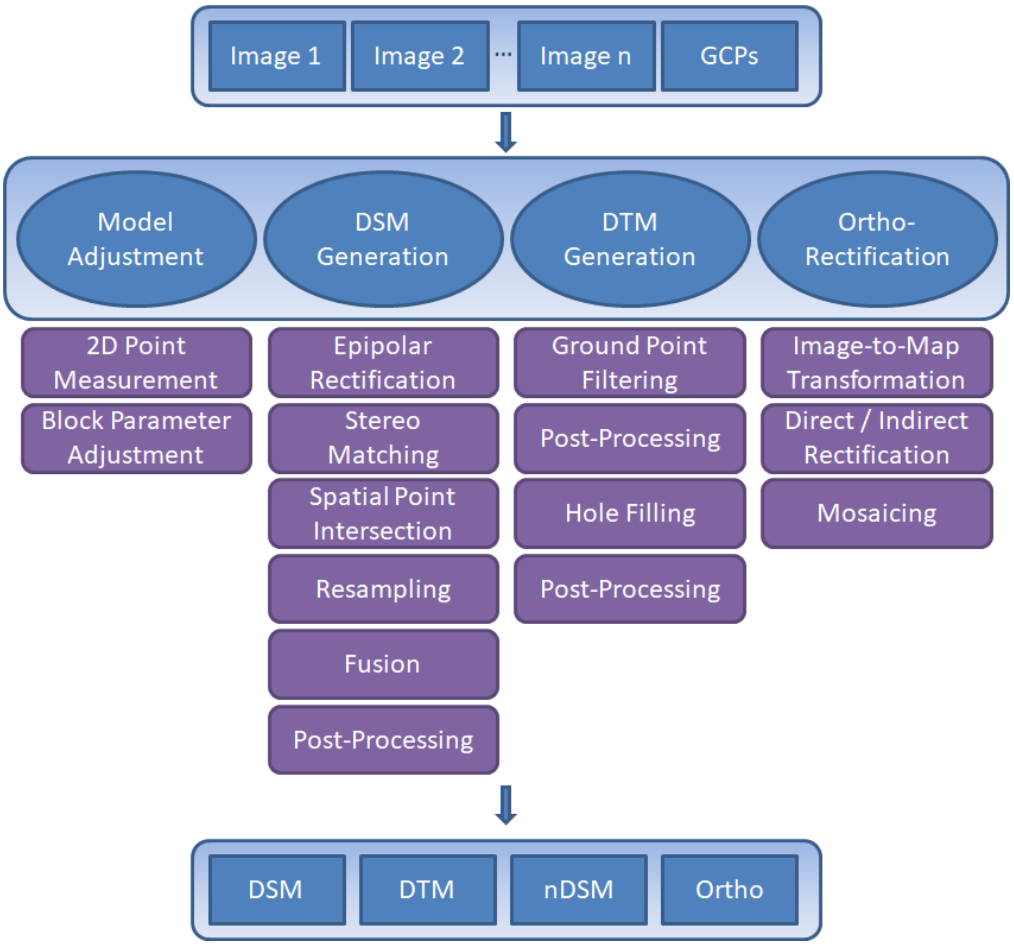


Figure 1.4: Proposed end-to-end 3D mapping workflow for multi-view stereo Pléiades images. The four core tasks are subdivided into according sub tasks.

- **Roland Perko**, Hannes Raggam, Mathias Schardt, and Peter M. Roth. "Very high resolution mapping with the Pléiades satellite constellation." *American Journal of Remote Sensing*, 6.2, pp. 89–99, December 2018 (Perko, Raggam, Schardt, et al., 2018).
- **Roland Perko**, Manuela Hirschmugl, Janik Deutscher, Mathias Schardt, Markus Hollaus, and Peter M. Roth. "Using multiple along and across track Pléiades stereo images for improved digital surface model generation." In *EARSeL Symposium*, Vol. 39, Salzburg, Austria, July 2019

Figure 1.5: Remote Sensing Software Graz: Selection of different suites. The topics of this objective diploma thesis are embedded within the Space Suite.

(Perko, Hirschmugl, Deutscher, et al., 2019).

- **Roland Perko**, Mathias Schardt, Livia Piermattei, Stefan Auer, and Peter M. Roth. "Multiple view geometry in remote sensing: An empirical study based on Pléiades satellite images." In *IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, July 2019 (Perko, Schardt, et al., 2019).
- **Roland Perko**, Hannes Raggam, and Peter M. Roth. "Mapping with Pléiades: End-to-End Workflow." *Remote Sensing*, under review, 2019 (Perko, Raggam, and Roth, 2019).

Closely related earlier publications are listed below.

- Karlheinz Gutjahr, **Roland Perko**, Hannes Raggam, and Mathias Schardt. "The epipolarity constraint in stereo-radargrammetric DEM generation." *IEEE Transactions on Geoscience and Remote Sensing*, 52.8, pp. 5014–5022, 52(8):5014-5022, 2014 (Gutjahr et al., 2014).
- **Roland Perko**, Hannes Raggam, Karlheinz Gutjahr, and Mathias Schardt. "Assessment of the mapping potential of Pléiades stereo and triplet data." In *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-3, pp. 103-109, Zurich, Switzerland, 2014 (Perko, Raggam, Gutjahr, et al., 2014).
- **Roland Perko**, Hannes Raggam, Karlheinz Gutjahr, and Mathias Schardt. "Advanced DTM generation from very high resolution satel-

lite stereo images." In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-3, pp. 165-172, Munich, Germany, 2015 (Perko, Raggam, Gutjahr, et al., 2015).

1.5 Outline

This thesis 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.

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 end-to-end mapping goal 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 1.4, 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 (Dial and Grodecki, 2002) 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, Triple-Sat, 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. (Grodecki and Dial, 2003)) 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 (Jacobsen and Topan, 2015; Fraser and Hanley, 2003) while linear coefficients are also considered in (Dial and Grodecki, 2002; Åstrand et al., 2012; M. Á. Aguilar, Mar Saldaña, and F. J. Aguilar, 2014; Perko, Raggam, Schardt, et al., 2018).

The need to manually measure GCPs and their according image coordinates represent the main bottleneck of mapping. Therefore, automatic methods for GCP definition and transfer are of utterly importance. One idea is to employ SAR images, like from the TerraSAR-X sensor. This sensor provides extremely high geo-location accuracy (Eineder et al., 2010) and thus perspectives to retrieve highly accurate 3D point information via radargrammetric processing (Raggam, Perko, et al., 2010; Raggam, Gutjahr, et al., 2010; Perko, Koppe, et al., 2018). Then such GCPs can be transferred into the optical Pléiades image by means of multi-modal image matching as presented, for instance, in (Perko, Raggam, Gutjahr, et al., 2011; Reinartz et al., 2011; Merkle et al., 2017).

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 1.4 the main steps are epipolar rectification, stereo matching, spatial point intersection, DSM resampling, DSM fusion, and according post-processing.

2.2.1 Epipolar Rectification

Epipolar rectification for push-broom images can be seen as a solved problem (M. Wang, Hu, and Li, 2011; Gutjahr et al., 2014). However, in contrast to perspective geometry, where analytic solutions exist for epipolar rectification (cf. (Fusiello, Trucco, and Verri, 2000; Fusiello and Irsara, 2008)), such rectification can only be approximated for push-broom images. The epipolar curves can be estimated by using the modified collinearity equations (Orun, 1994) and are hyperbola-like non-linear curves (Kim, 2000). 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, i.e., rectifying small parts of the images separately as shown in (De Franchis et al., 2014a; De Franchis et al., 2014b). 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 (Perko, Raggam, Schardt, et al., 2018). Recently, it was also shown that an epipolar rectification with small residual errors can be achieved for high resolution SAR satellite images (Gutjahr et al., 2014). Alternatively, epipolar rectification might be skipped if the subsequent stereo matching algorithm searches along the (non-linear) epipolar curves such as in (Hirschmüller, 2008). 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 (Scharstein and Szeliski, 2002). Having epipolar rectified images a standard procedure in photogrammetry and remote sensing is to apply the semi-global matching (SGM) introduced by (Hirschmüller, 2008). This matching scheme is able to achieve very good results, comparable to methods based on global optimization, e.g., graph cut (Kolmogorov and Zabih, 2001) or total variation (Ranftl, Pock, and Bischof, 2013). Moreover, it is computationally very efficient. Currently many commercial packages for processing satellite data use epipolar rectification and SGM for DSM generation, e.g., Erdas Imagine[1], PCI Geomatica[2], Trimble Inpho with Match-T or MATCH-3DX[3], s2p[4], SURE[5] (only works with perspective images (Rothermel et al., 2012)), MicMac[6] (Rupnik, Pierrot-Deseilligny, and Delorme, 2018), RPC Stereo Processor[7] (Qin, 2016), and also the Remote Sensing Software Graz[8]. Recent research reports on different variants of SGM, such as (Rothermel et al., 2012) and extensions like (Facciolo, De Franchis, and Meinhardt, 2015; Drory et al., 2014; Seki and Pollefeys, 2017; Scharstein, Taniai, and Sinha, 2017). As underlying cost function most algorithms apply the Hamming distance of the Census transform (Zabih and Woodfill, 1994) due to its robustness (Hirschmüller and Scharstein, 2007). Alternatively, the so-called modified Census Transform (Shin, Caudillo, and Klette, 2015; Klopschitz et al., 2017) is used. Also a weighting can be incorporated based on the radiometric or the geometric distance (similarity and proximity weighting) (Yoon and Kweon, 2006; Klopschitz et al., 2017). 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 en-

[1] http://www.hexagongeospatial.com/products/remote-sensing/erdas-imagine
[2] http://www.pcigeomatics.com
[3] http://www.trimble.com/imaging/inpho.aspx
[4] http://dev.ipol.im/~carlo/s2p/
[5] http://nframes.com/
[6] http://micmac.ensg.eu
[7] http://u.osu.edu/qin.324/rsp/
[8] http://www.remotesensing.at/en/remote-sensing-software.html

visaged but also the replacement of Census transform with generalized Census transform (Fife and Archibald, 2013; Ahlberg et al., 2019) is an issue. The most recent step is to train the cost-function via deep learning of a convolutional neural network (Zbontar and LeCun, 2016; Mayer et al., 2016; Zhu et al., 2017; Knöbelreiter, Vogel, and Pock, 2018).

2.2.3 Spatial Point Intersection

Using the sensor models and the disparities 3D points are extracted by means of 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 (Perko, Raggam, Schardt, et al., 2018).

2.2.4 DSM Resampling

There exist different paradigms, like storing the 3D information as point clouds (Nex et al., 2015; Schönfelder et al., 2017) 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 (Raggam, 2006). Another method is to extract DSMs from each stereo pair and then fuse those DSMs to a final DSM (Pock, Zebedin, and Bischof, 2011; Rumpler, Wendel, and Bischof, 2013; Perko and Zach, 2016). 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, e.g., interpolation of a regular grid of height values and filling the remaining

gaps by an appropriate interpolation mechanism (Perko, Raggam, Schardt, et al., 2018) or by LiDAR based approaches like (Pfeifer et al., 2014).

2.2.5 DSM Fusion

As mentioned above multiple DSMs stemming from individual stereo pairs (e.g., from a triple acquisition or from neighboring stereo acquisitions) have to be fused into one final DSM. One option is to apply local approaches like mean, median or probability mode based filtering (Rumpler, Wendel, and Bischof, 2013). The other option is to state a global optimization problem to be solved numerically via total variation (Pock, Zebedin, and Bischof, 2011) or via gradient descent (Perko and Zach, 2016).

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 (Davies, 2012; Hirschmüller, 2008). Small or isolated regions can then be discarded.

2.2.7 Additional Notes

Overall it can be stated that workflows exist that are able to 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[9]. The main challenge remains in finding the optimal parameters for image

[9]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 solution were benchmarked.

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)).

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 from spaceborne optical imagery.

2.3.1 Airborne LiDAR Data

There are many publications on DTM generation from LiDAR data, where first and last pulse data exist. Good review papers are (Sithole and Vosselman, 2004; Meng, Currit, and Zhao, 2010). 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. (Axelsson, 2000)). One representative algorithm is described in (Meng, L. Wang, et al., 2009), which will serve as the baseline for the algorithmic evolution presented in this thesis. A method on hierarchical robust linear prediction employing LiDAR data is introduced in (Kraus and Pfeifer, 2001).

2.3.2 Airborne Optical Data

For very high resolution airborne cameras the work (Wiechert and Gruber, 2010) reports on a method developed for UltraCam images (Leberl et al., 2003). 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 (Unger et al., 2009). 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 (Krauß, Arefi, and Reinartz, 2011). A standard method based on morphological opening was extended in (Krauß and Reinartz, 2010). 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 (Piltz, Bayer, and Poznanska, 2016). An automated stratified object-based approach that also uses the spectral information of an ortho image is presented in (Luethje, Tiede, and Eisank, 2017). An approach based on iterative geodesic reconstruction tested on Cartosat-1 stereo images is presented in (Arefi et al., 2011). DTM generation in forested regions based on Cartosat-1 stereo image is reported in (Tian,

Krauss, and Reinartz, 2014). There, the panchromatic image is classified into semantic classes, like buildings, low forest, high forest or ground, and different filtering is applied on 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 different 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 an map-to-image transformation. The process of mosaicing, i.e., 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 standard literature (Raggam, Almer, and Buchroithner, 1993; Toutin, 2011; Kraus, 2011; Campbell and Wynne, 2011; Förstner and Wrobel, 2016).

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 representative Pléiades acquisitions on the one hand, and appropriate reference data sets on the other hand, are available. The data from Singapore must not be published, thus no screenshots of this test site are presented in this thesis. Overall, 24 Pléiades images are used in this study, with acquisition dates and main acquisition parameters as listed in Table 3.1.

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 the 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 water 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.

Since the incidence angles of Pléiades images are given for across track (roll, ω), for along track (pitch, ϕ), and for the combined overall angle, we

derive equations to calculate the stereo intersection angle δ (also called the convergence angle (Jacobsen and Topan, 2015)) 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 δ and B/H as

$$r_1 = \begin{bmatrix} \tan\phi_1 \\ \tan\omega_1 \\ 1 \end{bmatrix}, \tag{3.1}$$

$$r_2 = \begin{bmatrix} \tan\phi_2 \\ \tan\omega_2 \\ 1 \end{bmatrix}, \tag{3.2}$$

$$\delta = \arccos\left(\frac{r_1 \cdot r_2}{|r_1| \cdot |r_2|}\right), \tag{3.3}$$

and

$$\frac{B}{H} = 2 \cdot \tan\frac{\delta}{2}. \tag{3.4}$$

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, i.e., eastern, center and western part. The central part of the test site is shown in Figure 3.1.

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 in 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

Table 3.1: Acquisition parameters for the Pléiades satellite images over the study areas.

Study Area	Acq. Date and Sensor	ID	Incidence Angle [°] along	across	overall	Model	Angle [°]
Innsbruck Stereo	2013-07-13 1A	1	-11.72	8.96	14.60	1-2	26.9
		2	14.48	2.35	14.66		
Innsbruck Triplet East	2014-06-09 1A	1	-18.74	4.11	19.13	1-2	10.1
		2	-8.99	1.42	9.09	2-3	23.3
		3	13.61	-4.63	14.32	1-3	33.4
Innsbruck Triplet Center	2014-06-08 1B	1	-17.82	-8.53	19.53	1-2	24.9
		2	6.54	-15.46	16.67	2-3	4.6
		3	11.15	-16.77	19.80	1-3	29.5
Innsbruck Triplet West	2014-06-09 1A	1	-14.30	3.82	14.76	1-2	10.4
		2	-4.21	1.13	4.36	2-3	23.0
		3	18.11	-4.89	18.67	1-3	33.4
Trento Triplet	2012-09-11 1A	1	-16.86	14.68	21.83	1-2	5.5
		2	-11.25	13.34	17.20	2-3	27.3
		3	15.74	7.00	17.09	1-3	32.9
Ljubljana Stereo 1	2013-07-27 1A	1	-12.12	-1.98	14.9	1-2	25.2
		2	12.36	-8.59	12.3		
Ljubljana Stereo 2	2013-07-28 1B	3	-7.39	9.36	15.3	3-4	22.7
		4	14.78	3.95	11.8		
Ljubljana Stereo 3	2013-07-29 1A	5	-10.20	22.31	24.1	5-6	26.2
		6	16.91	16.90	23.3		
Singapore Stereo	2014-10-01 1B	1	-0.31	8.64	8.64	1-2	15.2
		2	14.65	5.44	15.55		
Tian Shui Stereo	2014-05-07 1A	1	-10.54	0.75	10.56	1-2	22.8
		2	11.68	-4.54	12.48		

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.

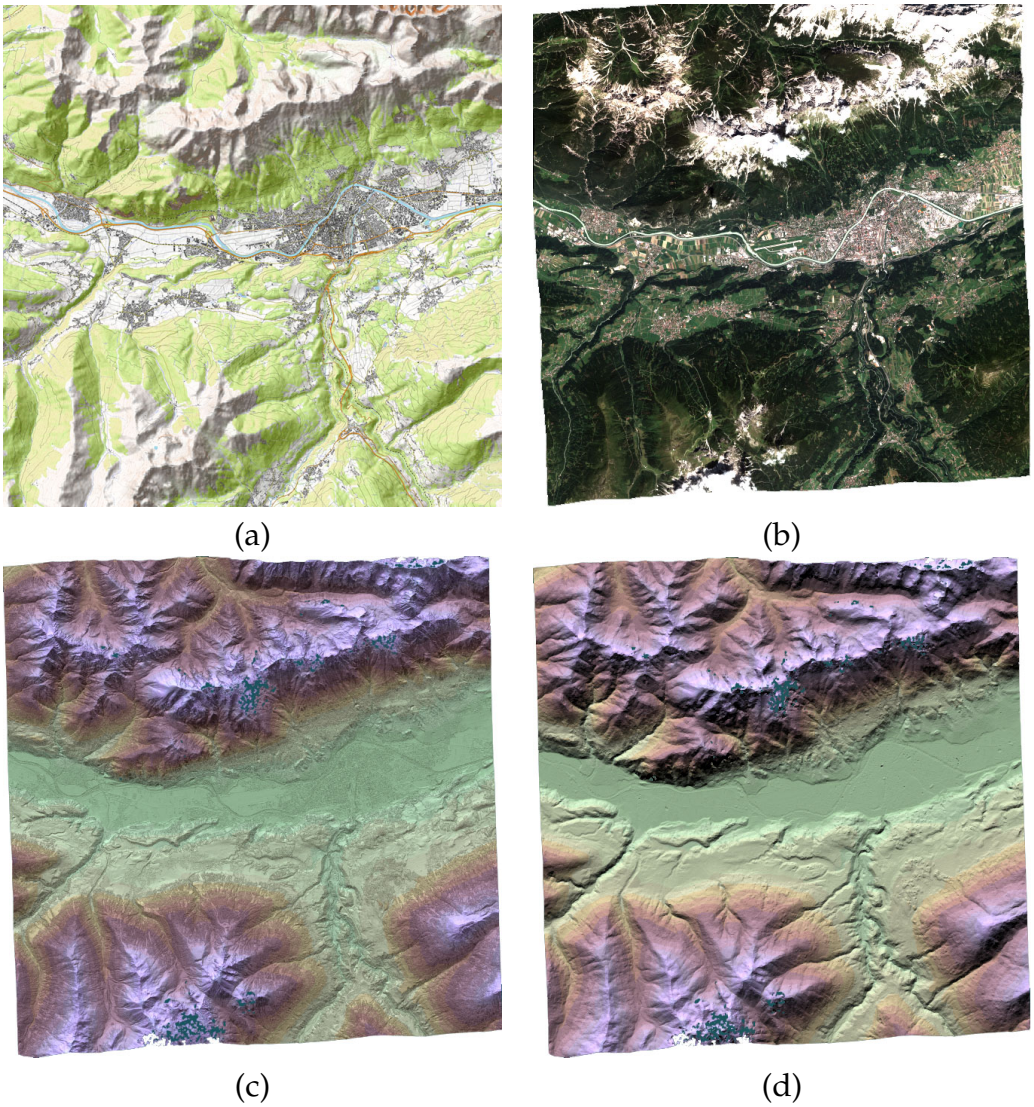


Figure 3.1: The test site Innsbruck (Austria): (a) topographic map (opentopomap.org (CC-BY-SA)), (b) RGB ortho-image, (c) relief shaded DSM, and (d) relief shaded DTM. (b-d) were produced employing the proposed and implemented mapping workflow.

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 3.2.

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.

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 (Piermattei, Marty, Karel, et al., 2018). 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.

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 (Pfeifer et al., 2014). 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

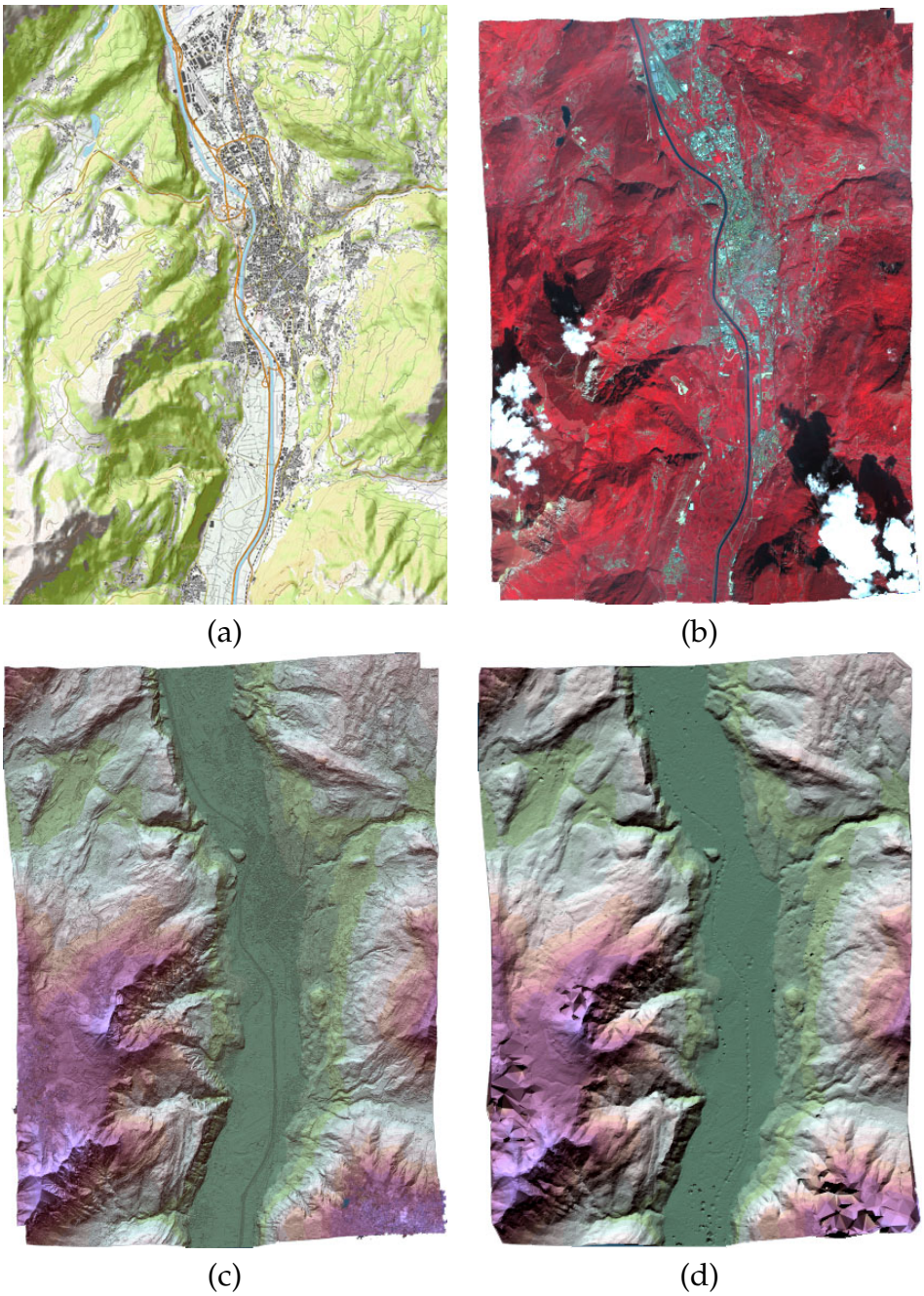


Figure 3.2: The test site Trento (Italy): (a) topographic map (opentopomap.org (CC-BY-SA)), (b) CIR ortho-image, (c) relief shaded DSM, and (d) relief shaded DTM. (b-d) were produced employing the proposed and implemented mapping workflow.

other data sets. For any other detail on the dataset we refer to (Piermattei, Marty, Karel, et al., 2018).

The acquisition scenario is depicted in Figure 3.3 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. 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 3.2 lists the convergence angles of all possible pairs based on the equations in Eq. (3.3). 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°. The test site is shown in Figure 3.4.

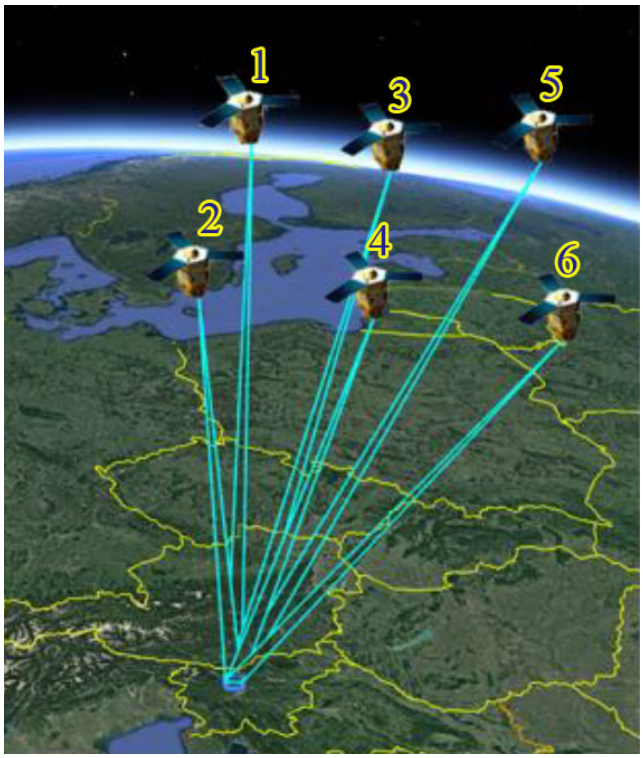


Figure 3.3: Multiple view geometry Ljubljana data set, adapted from (Piermattei, Marty, Karel, et al., 2018). The Google Earth visualization depicts the preview of the footprints and the satellite's position.

Table 3.2: 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 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.

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 (Leopold et al., 2017). 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 (TP) are

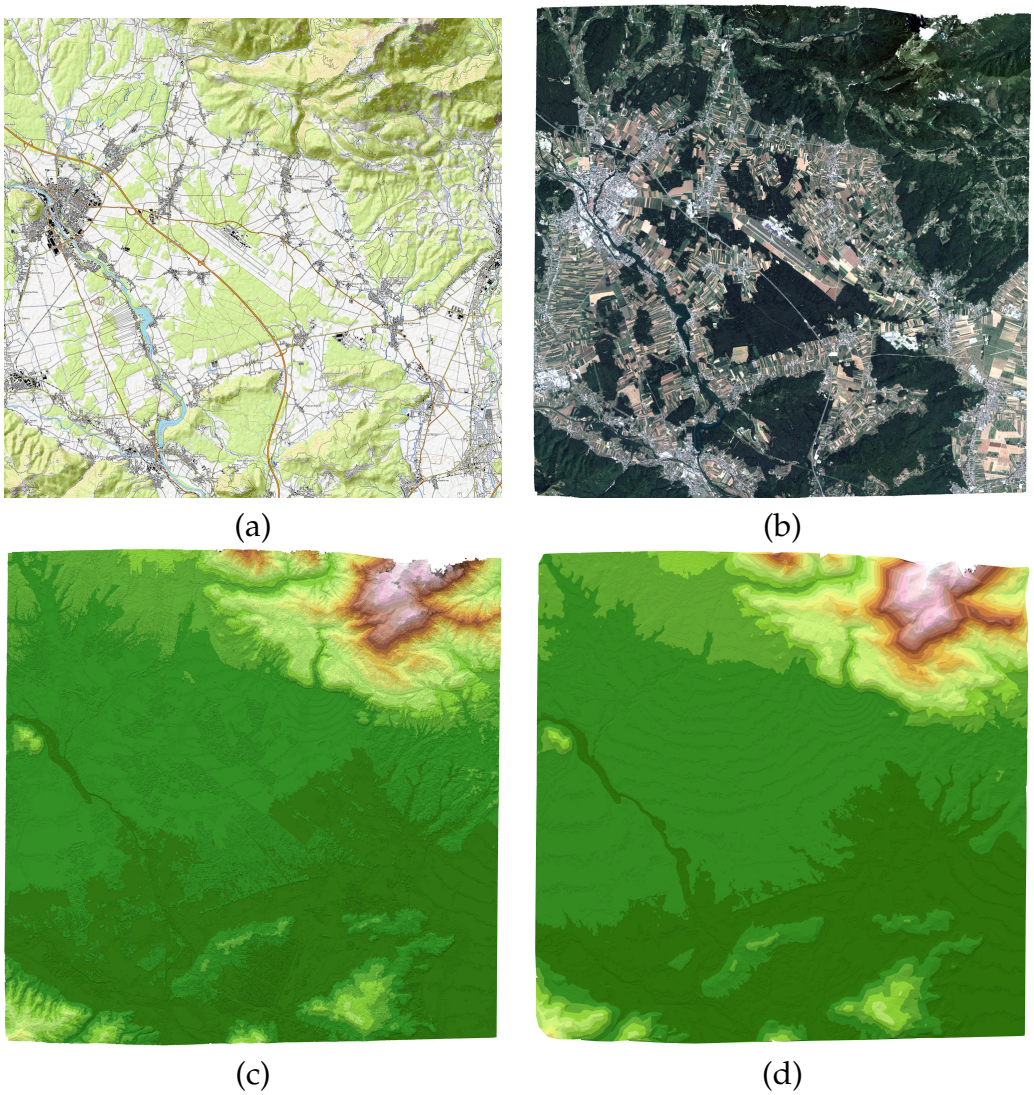


Figure 3.4: The test site Ljubljana (Slovenia): (a) topographic map (opentopomap.org (CC-BY-SA)), (b) RGB ortho-image, (c) relief shaded DSM, and (d) relief shaded DTM. (b-d) were produced employing the proposed and implemented mapping workflow.

manually measured and applied within the sensor model adjustment. The test site is shown in Figure 3.5.

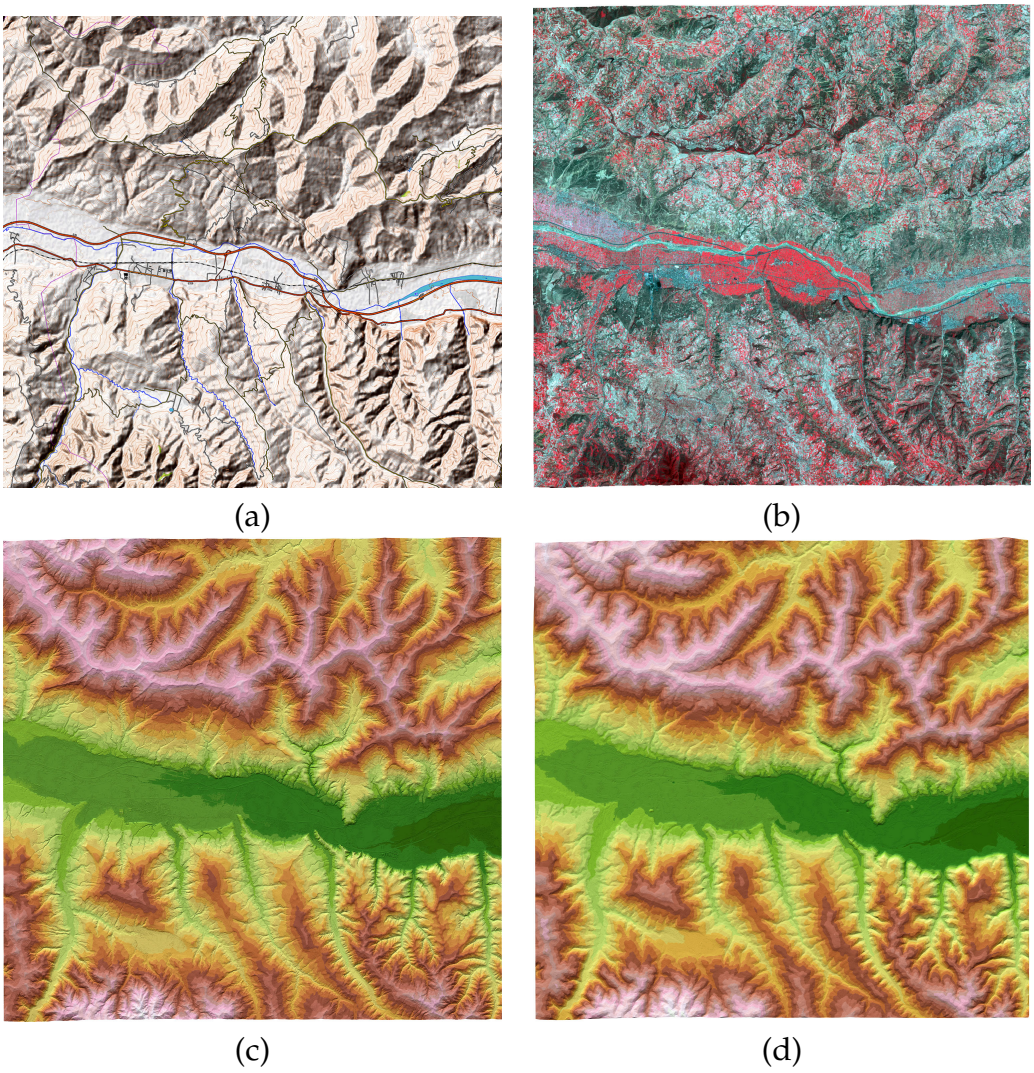


Figure 3.5: The test site Tian Shui (China): (a) topographic map (opentopomap.org (CC-BY-SA)), (b) CIR ortho-image, (c) relief shaded DSM, and (d) relief shaded DTM. (b-d) were produced employing the proposed and implemented mapping workflow.

4 Methods

The envisaged end-to-end mapping workflow comprises four core tasks to be efficiently accomplished (cf. Figure 1.4). Thus, in this section we discuss our algorithmic approaches with respect to (1) sensor modeling and parameter optimization, (2) generation of DSMs multi-stereo Pléiades images, (3) generation of DTMs using the extracted DSMs, and (4) ortho-rectification of the multi-spectral image data. The structure of this section follows Figure 1.4 and thus the related work presented in Section 2.

4.1 Sensor Modeling and Parameter Optimization

Since the Pléiades sensor model is delivered in form of RPCs, this model is described first. Then, the parameter optimization methodology is stated. Next, a protocol is defined that allows to assess the 2D and 3D geo-location accuracy of the sensor model.

4.1.1 The RPC Sensor Model

The RPC sensor model, as defined in (Dial and Grodecki, 2002; Grodecki and Dial, 2003), is briefly summarized here. We use the same variable names as in (Grodecki and Dial, 2003) for easier comparison. The RPC sensor model relates geographic coordinates (latitude ϕ, longitude λ, and height h) to corresponding image coordinates (line, sample / column). The RPC model itself is the ratio of two cubic polynomials of geographic (object-space) coordinates. Two rational functions are used to define the mapping from object-space to line, respectively, sample coordinates. Due to numerical reasons the object-space and image coordinates are normalized to $[-1, +1]$, by applying

offsets (LAT_OFF, LONG_OFF, HEIGHT_OFF, LINE_OFF, SAMP_OFF) and scale factors (LAT_SCALE, LONG_SCALE, HEIGHT_SCALE, LINE_SCALE, SAMP_SCALE). The normalization of object-space coordinates yields P, L, and H, and is calculated as follows:

$$P = \frac{\phi - LAT\_OFF}{LAT\_SCALE}, \tag{4.1}$$

$$L = \frac{\lambda - LONG\_OFF}{LONG\_SCALE}, \tag{4.2}$$

$$H = \frac{h - HEIGHT\_OFF}{HEIGHT\_SCALE}. \tag{4.3}$$

The normalized line and sample image coordinates (Y and X) are then calculated via the rational polynomial functions

$$Y(P,L,H) = \frac{Num_L(P,L,H)}{Den_L(P,L,H)} = \frac{\boldsymbol{c}^T\boldsymbol{u}}{\boldsymbol{d}^T\boldsymbol{u}} \tag{4.4}$$

and

$$X(P,L,H) = \frac{Num_S(P,L,H)}{Den_S(P,L,H)} = \frac{\boldsymbol{e}^T\boldsymbol{u}}{\boldsymbol{f}^T\boldsymbol{u}} \tag{4.5}$$

with

$$Num_L(P,L,H) = \sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3} c_{ijk}P^iL^jH^k, \tag{4.6}$$

$$Den_L(P,L,H) = \sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3} d_{ijk}P^iL^jH^k, \tag{4.7}$$

$$Num_S(P,L,H) = \sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3} e_{ijk}P^iL^jH^k, \tag{4.8}$$

and

$$Den_S(P,L,H) = \sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3} f_{ijk}P^iL^jH^k. \tag{4.9}$$

Furthermore, the total power of all three ground coordinates it limited to 3, i.e., the coefficients are zero whenever $i + j + k > 3$. Following this approach the image coordinates (Y and X) can be written as fractions of vector dot products (cf. Eq. (4.4) and Eq. (4.5)) as

$$\begin{aligned} \boldsymbol{u} &= [1, L, P, H, LP, LH, PH, L^2, P^2, H^2, PLH, && (4.10)\\ & \quad L^3, LP^2, LH^2, L^2P, P^3, PH^2, L^2H, P^2H, H^3]^T\,, && (4.11)\end{aligned}$$

$$\boldsymbol{c} = [c_1, c_2, \ldots, c_{20}]^T\,, \tag{4.12}$$

$$\boldsymbol{d} = [1, d_2, \ldots, d_{20}]^T\,, \tag{4.13}$$

$$\boldsymbol{e} = [e_1, e_2, \ldots, e_{20}]^T\,, \tag{4.14}$$

and

$$\boldsymbol{f} = [1, f_2, \ldots, f_{20}]^T\,. \tag{4.15}$$

Finally, the de-normalizing of image coordinates is determined as follows:

$$Line = Y \cdot LINE\_SCALE + LINE\_OFF \tag{4.16}$$

and

$$Sample = X \cdot SAMP\_SCALE + SAMP\_OFF\,. \tag{4.17}$$

Overall, the sensor RPC model consists of 80 coefficients (where d_1 and f_1 are set to 1) plus the 10 offset and scale values and approximates the physical sensor model in a standardized manner.

4.1.2 Parameter Optimization

For optimizing an RPC sensor model GCPs and their according image coordinates are employed. The optimization can be formulated in object-space or in image-space (Grodecki and Dial, 2003). 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 (Dial and Grodecki, 2002; Jacobsen and Topan, 2015). From the mathematical perspective the parameter optimization can be formulated as a non-linear equation system of the form

$$F(\boldsymbol{x}) = \boldsymbol{0}\,. \tag{4.18}$$

Actually our system consists of Y in Eq. (4.4) and X in Eq. (4.5). To solve this system we use Newton's method (described in Algorithm 4.1) with the starting vector $\boldsymbol{x}_0$. Via linearization we get

$$F(\boldsymbol{x} + \Delta\boldsymbol{x}) \approx F(\boldsymbol{x}) + J_F(\boldsymbol{x})\Delta\boldsymbol{x} \tag{4.19}$$

with the Jacobian matrix

$$J_F(\boldsymbol{a}) := \frac{\partial F}{\partial \boldsymbol{x}}(\boldsymbol{a}) = \left(\frac{\partial F_i}{\partial x_j}(\boldsymbol{a})\right)_{i,j} = \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(\boldsymbol{a}) & \frac{\partial F_1}{\partial x_2}(\boldsymbol{a}) & \cdots & \frac{\partial F_1}{\partial x_n}(\boldsymbol{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1}(\boldsymbol{a}) & \frac{\partial F_m}{\partial x_2}(\boldsymbol{a}) & \cdots & \frac{\partial F_m}{\partial x_n}(\boldsymbol{a}) \end{bmatrix}. \tag{4.20}$$

Algorithm 4.1: Solving non-linear equation systems with Newton's method.

Input:
non-linear equation system of form $F(\boldsymbol{x}) = \boldsymbol{0}$
and its Jacobian matrix J_F
starting vector $\boldsymbol{x}_0$
maximal iterations `// E.g., set to 20.`
tolerance `// E.g., set to 1e−7.`
Output:
solution vector $\boldsymbol{x}_{n+1}$

Function `NewtonsMethod`(F, J_F, $\boldsymbol{x}_0$, iterations, tolerance)**:**
  for $n = 0 : iterations$ **do**
    $J_F(\boldsymbol{x}_n)\Delta\boldsymbol{x}_n + F(\boldsymbol{x}_n) = \boldsymbol{0}$ `// Solve for` $\Delta\boldsymbol{x}_n$ `via least squares.`
    $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \Delta\boldsymbol{x}_n$ `// Get next approximation.`
    if $(|\Delta\boldsymbol{x}_n| \leq tolerance \cdot |\boldsymbol{x}_n|)$ **then**
      `break` `// Solution found within given tolerance.`
    end
  end
  return $\boldsymbol{x}_{n+1}$ `// Return solution vector.`

The linearization for the RPC model can now be formalized starting with the partial derivatives of $\boldsymbol{u}$ (cf. Eq. (4.11)) as

$$\frac{\partial \boldsymbol{u}}{\partial P} = [0, 0, 1, 0, L, 0, H, 0, 2P, 0, LH, 0, 2LP, 0, L^2, 3P^2, H^2, 0, 2PH, 0]^T \tag{4.21}$$

$$\frac{\partial \boldsymbol{u}}{\partial L} = [0, 1, 0, 0, P, H, 0, 2L, 0, 0, PH, 3L^2, P^2, H^2, 2LP, 0, 0, 2LH, 0, 0]^T \quad (4.22)$$

$$\frac{\partial \boldsymbol{u}}{\partial H} = [0, 0, 0, 1, 0, L, P, 0, 0, 2H, PL, 0, 0, 2LH, 0, 0, 2PH, L^2, P^2, 3H^2]^T \quad (4.23)$$

such that

$$\nabla \boldsymbol{u} = \left(\frac{\partial \boldsymbol{u}}{\partial P}, \frac{\partial \boldsymbol{u}}{\partial L}, \frac{\partial \boldsymbol{u}}{\partial H} \right), \quad (4.24)$$

which defines our partial derivatives of Y w.r.t. P, L, and H and thus the first line of our Jacobian matrix (cf. Eq. (4.20)) as

$$\frac{\partial Y}{\partial P} = \frac{\boldsymbol{c}^T \frac{\partial \boldsymbol{u}}{\partial P} \boldsymbol{d}^T \boldsymbol{u} - \boldsymbol{d}^T \frac{\partial \boldsymbol{u}}{\partial P} \boldsymbol{c}^T \boldsymbol{u}}{(\boldsymbol{d}^T \boldsymbol{u})^2}, \quad (4.25)$$

$$\frac{\partial Y}{\partial L} = \frac{\boldsymbol{c}^T \frac{\partial \boldsymbol{u}}{\partial L} \boldsymbol{d}^T \boldsymbol{u} - \boldsymbol{d}^T \frac{\partial \boldsymbol{u}}{\partial L} \boldsymbol{c}^T \boldsymbol{u}}{(\boldsymbol{d}^T \boldsymbol{u})^2}, \quad (4.26)$$

and

$$\frac{\partial Y}{\partial H} = \frac{\boldsymbol{c}^T \frac{\partial \boldsymbol{u}}{\partial H} \boldsymbol{d}^T \boldsymbol{u} - \boldsymbol{d}^T \frac{\partial \boldsymbol{u}}{\partial H} \boldsymbol{c}^T \boldsymbol{u}}{(\boldsymbol{d}^T \boldsymbol{u})^2}. \quad (4.27)$$

The linearization for the line coordinates Y (cf. Eq. (4.4) and Eq. (4.19)) can now be written as

$$Y(P + \Delta P, L + \Delta L, H + \Delta H) \approx Y(P, L, H) + \frac{\partial Y}{\partial P}\Delta P + \frac{\partial Y}{\partial L}\Delta L + \frac{\partial Y}{\partial H}\Delta H. \quad (4.28)$$

The linearization for the sample coordinates X can be formalized accordingly and they define the second line of our Jacobian matrix (cf. Eq. (4.20)). For each GCP in one image we get two equations.

The goal in parameter optimization can be seen as to optimize as few parameters as reasonable for two reasons: (1) to reduce the need for GCPs and (2) to avoid drifting or even oscillating of the RPCs. Thus, two options of optimizations are presented in this work. The first only models a shift, i.e., releasing the parameters c_1 and e_1 of the RPC nominators. In theory only one single GCP is needed for such an adjustment. The second releases constant and linear terms of the RPC nominators, i.e., c_i and e_i with $i \in [1, 4]$. While a minimum of only four GCPs would be sufficient to optimized those

eight parameters, over-determination is reasonable in order to mitigate the impact of measurement errors. The decision to adjust the linear nominator coefficients is based on previous experiments involving different adjustment settings, which have shown that in this way sub-pixel accuracy can be achieved. These findings are also confirmed by previous work (Dial and Grodecki, 2002; Åstrand et al., 2012; M. Á. Aguilar, Mar Saldaña, and F. J. Aguilar, 2014). Releasing higher order terms the RPC model usually gets drawn towards the GCPs which results in overfitting. This is also the reason why we use GCPs and ICPs in the evaluation. The GCPs are used to adjust the sensor model, while the ICPs are used to evaluate the resulting model. Therefore, a potential overfitting can be detected and thus avoided.

4.1.3 Geo-Location Accuracy Assessment

To assess the geo-location accuracies of the initial and optimized sensor models, reference GCPs and also their 2D locations in the images serve as input data. Then, the 3D GCPs are projected into the 2D image using the RPC model and the deviations to the reference 2D image coordinates define the model accuracy. Those differences are described as statistical values, like 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 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 Figure 1.4 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 the approach in (M. Wang, Hu, and Li, 2011) is applied. As described in (Gutjahr et al., 2014) the main steps are as follows. First, both stereo images are projected onto a common reference plane, which is defined via the mean height of the mapped area. This step already ensures that the epipolar curves can be assumed as parallel lines with deviations below the image resolution. In a second step, a 2D Helmert transformation with a fixed scale is applied to both projected images. The rotation angle of this transformation can be calculated as described in (M. Wang, Hu, and Li, 2011), and thus, the transformation ensures that the direction of the epipolar lines corresponds to the column direction of the images. The shifts of this transformation ensure zero parallaxes in the line direction and control the parallaxes in the column direction. For the calculation of the shift parameters, we transform the center point of the warped left scene to the warped right scene using the mean height of the given scene. To avoid repetitive resampling, both processing steps are combined, thus establishing the transformation from a pixel in the epipolar image to the corresponding position in the original image (i.e., indirect resampling). This can be done on a pixel-by-pixel basis independently from the resampling which is done in a final step. As resampling method the 6-point cubic convolution (Keys, 1981) is used, as it yields highly accurate results while being computationally feasible (higher order resampling methods or windowed Sinc approaches are computationally more demanding (Meijering, Niessen, and Viergever, 2001; Perko, 2004)).

Since, the transformations are non-linear and also incorporate iterative image-to-map transforms (Newton's method and linearization), pixel-by-pixel evaluation is computationally expensive. Especially, as the epipolar Pléiades images are huge, with around 50000×50000 pixels for a full scene,

an approximation is desired. It can be observed that the transformation is a locally slowly changing continuous function. Thus, an evaluation on grid points only is sufficient. In between the grid cells bilinear interpolation can be used. Empirical tests show that a grid step size of 100 pixels can be used without losing accuracy, thus reducing the transformation effort by a factor of 10000. 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 4.1 depicts the 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). Overall, the proposed approximation results in a significant speedup.

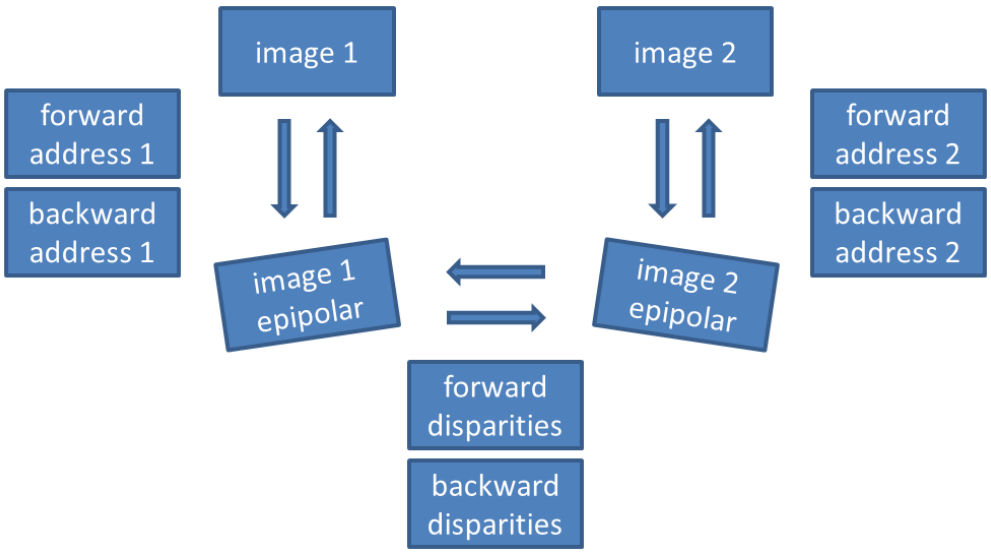


Figure 4.1: Relation of stereo input images, their epipolar rectified versions, forward and backward addresses, and both disparity maps.

4.2.2 Image Matching

As the image matching is employed on epipolar images, the disparities have to be searched along one dimension only, i.e., the column direction. The classical matching paradigm is depicted in Figure 4.2 and contains

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).

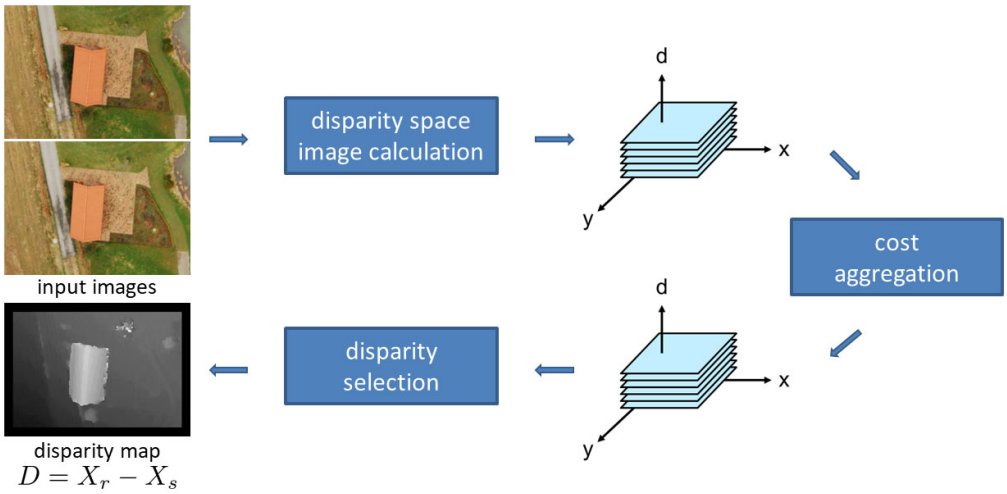


Figure 4.2: Image matching principle with epipolar rectified input images. The main steps are calculation of the disparity space image, cost aggregation, and disparity selection. The final disparity map D holds the shifts in column directions for each pixel such that they point from the reference image X_r to the search image X_s.

The current state-of-the-art in cost volume calculation is the normalized Hamming distance of the Census transform (Zabih and Woodfill, 1994), as it is invariant to non-linear radiometric distortions and very efficient in its calculation. The Census transform $\xi(X, x_c)$ is defined for an image patch X with pixels x_i and the central pixel value x_c as

$$\xi(X, x_c) = \begin{cases} 0 & x_i \leq x_c \\ 1 & x_i > x_c \end{cases} \tag{4.29}$$

and the according cost-function $C_\xi(X, Y)$ that compares two patches X and Y, where n is the number of pixels in X, and Hamming the Hamming distance as

$$C_\xi(X, Y) = \frac{1}{n}\text{Hamming}(\xi(X, x_c), \xi(Y, y_c)) \in [0, 1]\,. \tag{4.30}$$

As this definition compares all pixels within a patch X to the central pixel value x_c the transform becomes ambiguous when the patch holds homogeneous, i.e., textureless, values. Therefore, the modified Census can be used that compares the pixels to the mean value of the patch (Hermann and Klette, 2009; Klopschitz et al., 2017). Another option that works well for satellite based images is to stabilize the Census transform by combination with a *sum of absolute differences* (SAD) cost function $C_\epsilon(X, Y)$, with

$$C_\epsilon(X, Y) = \frac{1}{n \cdot m} \sum_i |x_i - y_i| \in [0, 1] \,, \tag{4.31}$$

where m denotes the scaling factor that depends on the pixel depth of the current image. For 8 bit images, we use $m = 2^8$ and for 16 bit images, like it is the case for Pléiades, we use $m = 2^{16}$. The combined cost-function is then defined as

$$C_{\xi,\epsilon}(X, Y) = \frac{1}{2} \left(C_\xi(X, Y) + C_\epsilon(X, Y) \right) \in [0, 1] \,. \tag{4.32}$$

Normalized cross correlation yields more blurring at 3D breaklines than Census transform such that it is not applied (cf. (Hirschmüller and Scharstein, 2007)). 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 (Bleyer and Chambon, 2010).

After calculation of the cost volume the cost aggregation is performed with semi-global matching (SGM) introduced in (Hirschmüller, 2008). 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. (Scharstein, Taniai, and Sinha, 2017)). 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 (Drory et al., 2014). Overall, SGM approximates the energy minimization of a 2D Markov random field

$$E(D) = \sum_{p} C_p(d_p) + \sum_{p,q \in \mathcal{N}} V(d_p, d_q)\,, \tag{4.33}$$

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

$$V(d, d') = \begin{cases} 0 & \text{if } d = d' \\ P_1 & \text{if } |d - d'| = 1 \\ P_2 & \text{if } |d - d'| \geq 2\,, \end{cases} \tag{4.34}$$

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 Eq. (4.33) along 8 cardinal directions r via dynamic programming (Hirschmüller, 2008). For each direction r an aggregated matching cost $L_r(p, d)$ is recursively calculated as

$$L_r(p, d) = C_p(d) + \min_{d' \in \mathcal{D}} \left(L_r(p - r, d') + V(d, d') \right). \tag{4.35}$$

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

$$S(p, d) = \sum_{r} L_r(p, d)\,, \tag{4.36}$$

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

$$d_p = \arg\min_{d} S(p, d)\,. \tag{4.37}$$

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 r (Hirschmüller, 2008). 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, r, and three parameters ϵ, δ, and P_2' as

$$P_2(I, \epsilon, \delta) = \max\left(\frac{\mu_I \cdot P_2'}{|I_p - I_q| + \epsilon}, P_1 + \delta \right) \tag{4.38}$$

with μ_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 ϵ to avoid a division by zero if I_p equals I_q. The parameter ϵ is set to $\epsilon = \{\min(|I_p - I_q|) \mid \forall |I_p - I_q| \neq 0\}$, i.e., ϵ 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 δ. We set δ in a pragmatic way to $\delta = P_1 + (P_2^{'} - P_1)/100$ 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 matching algorithm.

The resulting disparities d_p are calculated at integer precision. To get subpixel refinements the classical approach is to fit a polynomial of order 2, i.e., 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 3×3 window, so for instance 1.5 pixel. Then also a threshold on the aggregated cost is performed. Lastly a peak filter detects connected regions (cf. (Hirschmüller, 2008)), where small islands are discarded. In the next step, discarded pixels are classified as blunder or as occlusions as described in (Hirschmüller, 2008). 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.

We extend the algorithm mainly to gain a significant speed-up as follows. While in (Hirschmüller, 2008) 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 (Paar and Pölzleitner, 1992) to limit the search range for each pyramid level (visualized

in Figure 4.3). This 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., 15 pixels instead of 1550 pixels, which is the average disparity range of the Ljubljana data set). We call this method *truncated SGM*. As depicted in Figure 4.3 it might be the case that for a disparity d the adjacent disparity d' is not calculated thus does not exist. Therefore, the smoothness term is changed to

$$V(d,d') = \begin{cases} \text{if } d' \text{ exists} & \begin{cases} 0 & \text{if } d = d' \\ P_1 & \text{if } |d-d'| = 1 \\ P_2 & \text{if } |d-d'| \geq 2 \end{cases} \\ \text{else} & P_2\,. \end{cases} \tag{4.39}$$

Depending on the 3D structure of the scene maximal memory consumption reduction between 65% to 95% is achieved while the runtime is reduced by 30% to 90% (also cf. (Rothermel et al., 2012)). 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% of memory reduction. 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 by reduced by around 33% (mean value over our test images).

To gain speedup and accuracy, we incorporate a coarse digital elevation model (DEM) (e.g., SRTM DEM (Farr et al., 2007), ASTER GDEM (Tachikawa et al., 2011), WorldDEM (Riegler, Hennig, and Weber, 2015)) 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.

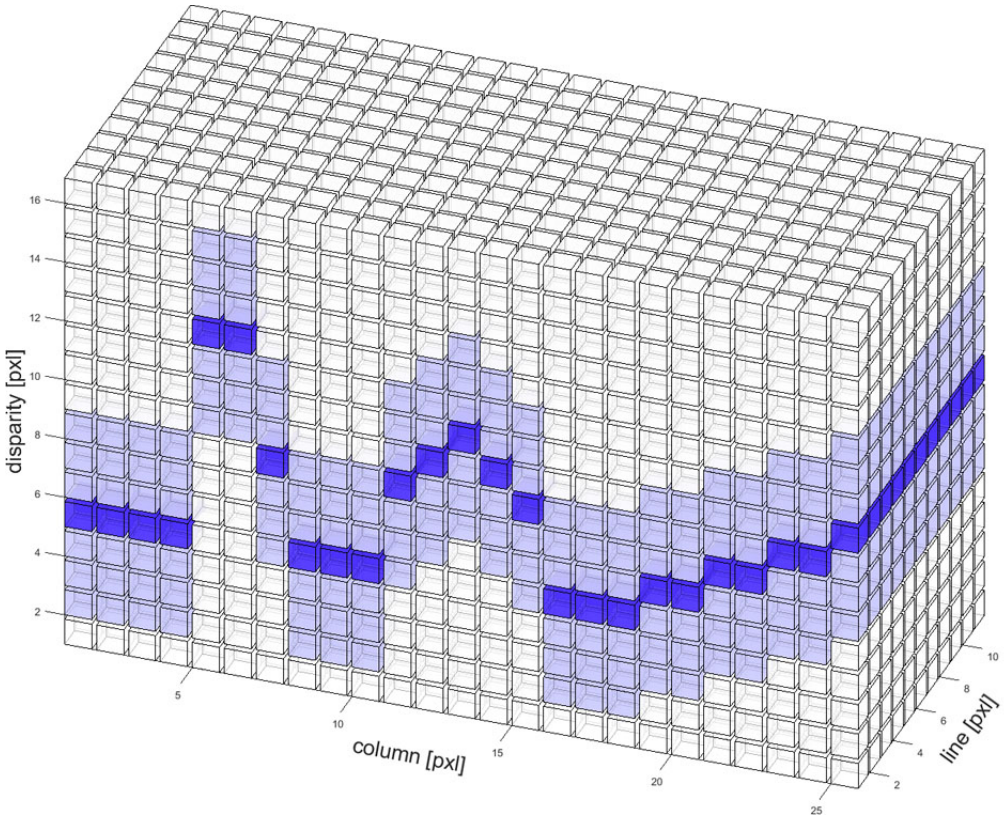


Figure 4.3: Cost volume $C_p(d)$ of classical SGM versus our implementation is shown on a toy example. While SGM evaluates the whole cost volume, we only check the disparities visualized in light blue color around the disparity predictions from the previous pyramid level depicted in dark blue color. In this toy example the search space is reduced to 44%, however, the reduction is larger in real satellite images.

Figure 4.1). 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.

As others, we found that the simple parabola subpixel interpolation yields a bias in the distribution of fractional disparities (cf. (Haller and Nedevschi, 2012)). 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 4.4.

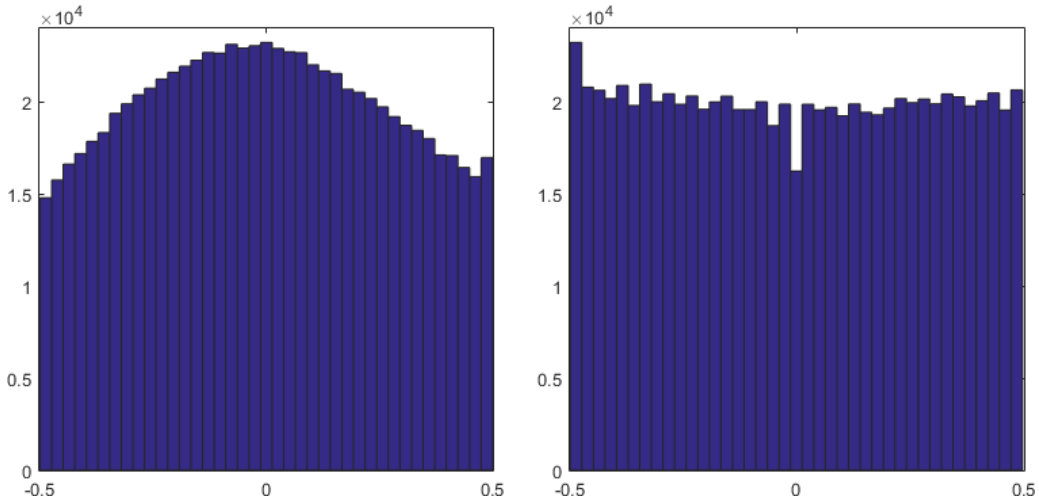


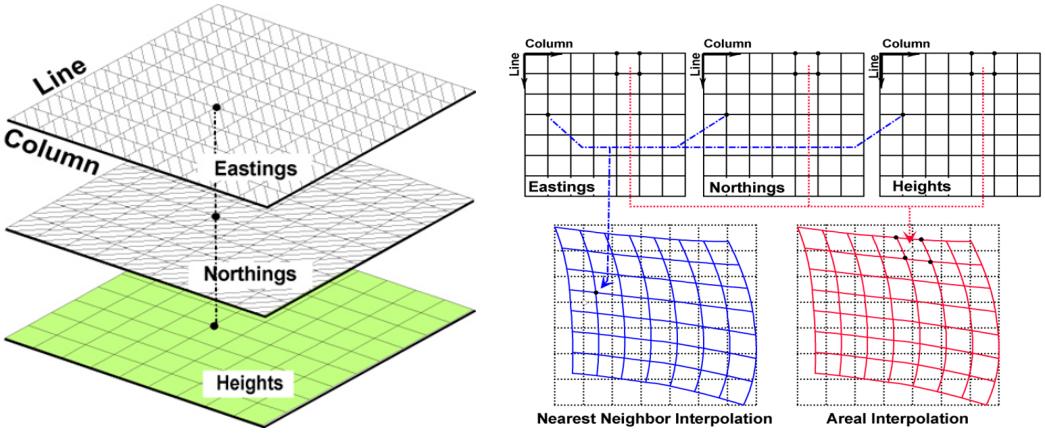
Figure 4.4: Subpixel distribution (a) based on the parabola interpolation (Hirschmüller, 2008) and (b) after our proposed normalization.

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 4.1 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 (Isenburg, 2013) or in a multi-band image. In this thesis 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 4.5 (a) and (Schönfelder et al., 2017)).

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 for doing so. First, the dimension and spacing of the output DSM is defined. Then, we use either nearest neighbor interpolation or areal interpolation in the regridding step. In the firmer 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 4.5.



(a) (b)

Figure 4.5: 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 (Rumpler, Wendel, and Bischof, 2013) which takes all height measurements within a 3×3 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 extent 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 (Rumpler, Wendel, and Bischof, 2013). Alternative fusion methods based on global optimizations like (Perko and Zach, 2016; Pock, Zebedin, and Bischof, 2011) would be of interest as well but are neglected in this study due to their high computational complexity.

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 $\mathrm{DSM}_{\Delta} = \mathrm{DSM}_{\mathrm{ref}} - \mathrm{DSM}_{\mathrm{current}}$, 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 statistics 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. (Rousseeuw and Croux, 1993; J. Höhle and M. Höhle, 2009; Leys et al., 2013; Jacobsen and Topan, 2015)). 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

$$\mathrm{NMAD}(\Delta H) = 1.4826 \cdot \mathrm{median}(|\Delta H - \mathrm{median}\,\Delta H|)\,. \tag{4.40}$$

Figure 4.6 depicts two real examples of this study where only errors between -10 m and $+10$ m are plotted. Shown are 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 both plots 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. In both plots 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.

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 1.4, 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.

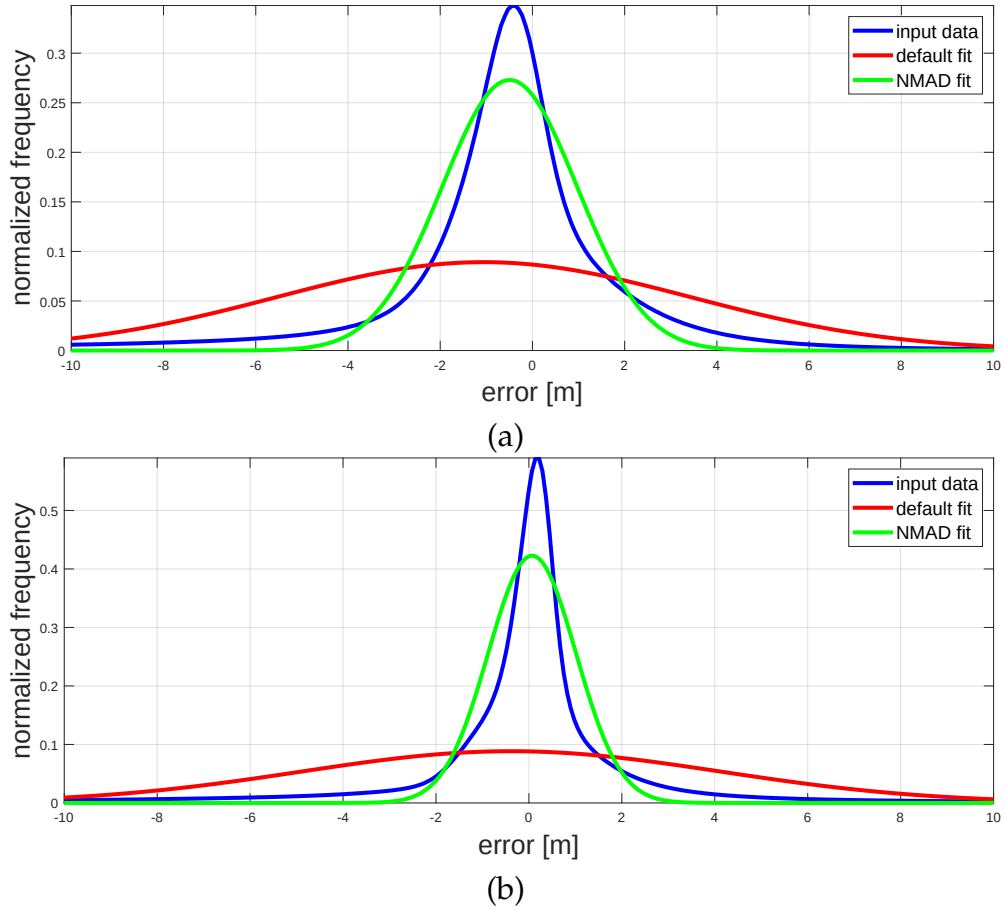


Figure 4.6: Two examples of fitting a normal distribution into non-normal distributed input data (blue). Default fit based on mean and standard deviation estimates (red) and robust fit based on median and NMAD estimates (green).

4.3.1 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 (Meng, L. Wang, et al., 2009), which is designed for LiDAR data,

to be slope dependent and to be really multi-directional, i.e., 8-directional to span the complete 2D image space. Nonetheless, focus is put onto simplicity, robustness and computational efficiency to follow integration needs into the automatic end-to-end workflow. Overall, the main concepts of DSM to DTM filtering can be summarized as follows:

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

Obviously, the crucial and most difficult step is the first one. Based upon our literature study the most promising algorithm to start with is the directional filtering method by (Meng, L. Wang, et al., 2009; Meng, Currit, and Zhao, 2010). A short recap on directional filtering according to (Meng, L. Wang, et al., 2009) can be given as follows: The basic idea is to process each line of the given DSM separately, e.g., from left to right with a sliding window of 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 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 method. The algorithm uses three parameters, i.e., 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 (Meng, L. Wang, et al., 2009; Meng, Currit, and Zhao, 2010) 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. This issue is sketched in Figure 4.7 where a 1D profile of an artificial DSM to be filtered is given in (a). It shows a tilted surface with some noise and a building. 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 95.8 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 (99.2 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 4.7.

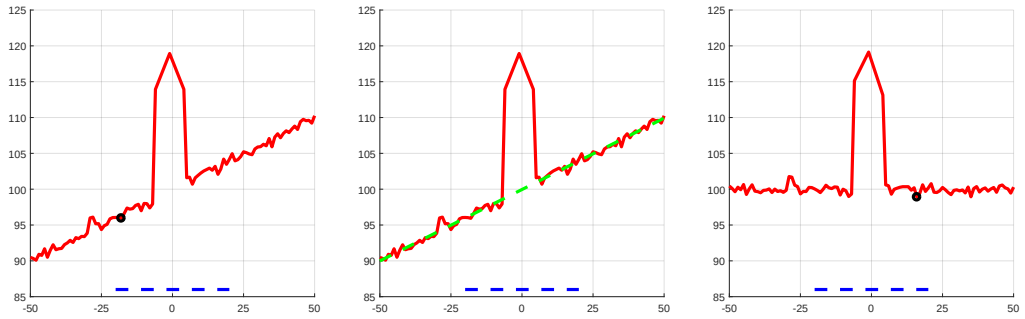
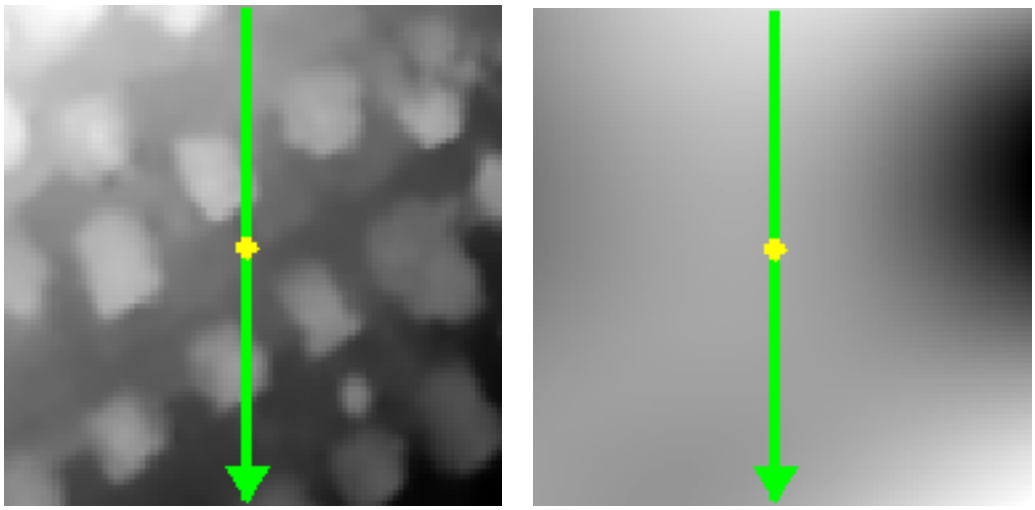


Figure 4.7: DSM profiles for an artificial building on a tilted surface with added noise: (a) Original DSM profile and the minimal height value (black) within the window (blue). (b) Robust fit of the surface (green). (c) Slope corrected DSM profile and the according minimal height value.

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. (Fox and Weisberg, 2018)). However, this robust fit has to be performed for each pixel and for all scanline directions and thus would slow down the process. Experiments have shown that a simple fit via smoothing with a huge kernel size yields very similar and thus satisfactory results. Therefore, a 2D Gaussian smoothing is implemented with a spatial sigma σ 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.

Figure 4.8 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 in Figure 4.9 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.



(a) (b)

Figure 4.8: Proposed simple fit on the terrain slope: (a) shows a subset of a DSM and (b) its Gaussian filtered variant, representing the slope of the profile.

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 (Hirschmüller, 2008). 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 4.2 and Algorithm 4.3 resulting in a ground and a non-ground label image. Instead of

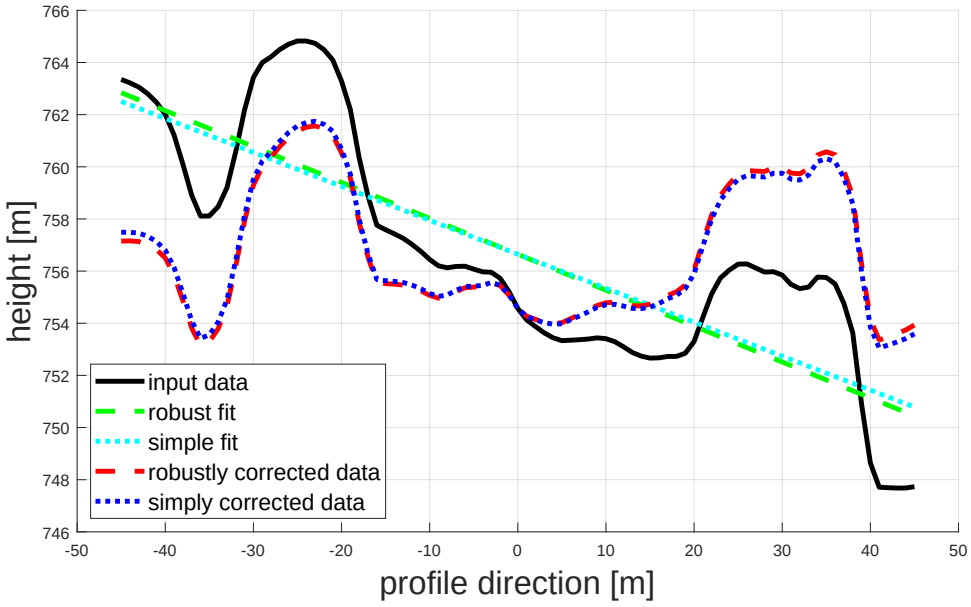


Figure 4.9: Comparison of robust terrain slope fit to the proposed simple fit. Shown are the profile of the DSM along the green arrow in Figure 4.8 together with the terrain slope fits and the corrected profiles.

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 4.10). 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 a 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.

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.

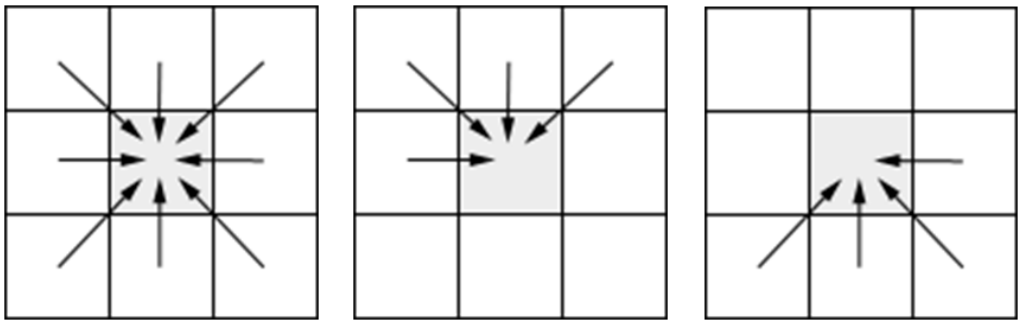


Figure 4.10: Concept of splitting 8 scanlines spanning the 2D image space (a) into 2 passes (b) down-right and (c) up-left.

- 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 (Stein, 2012), anisotropic diffusion (Weickert and Welk, 2006), total variation (Unger et al., 2009), or total generalized variation (Bredies, Kunisch, and Pock, 2010).

4.4 Ortho-Rectification

Standard indirect 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.

Algorithm 4.2: Multi-directional slope dependent DTM generation (part 1).

Input:

DSM // Digital surface model.
Spacing // Spacing of DSM in meters.
DSMsmooth // Smoothed DSM.
Dir // Scanline direction represented as two // integer shifts ShiftX and ShiftY.
ExtendMetric // Filter extend in meters, converted to // pixels using Spacing and yielding Ext.
Ext // Filter extend (odd) which also defines the local // direction dependent neighborhood // ExtX(Dir) and ExtY(Dir).
X // centered number line with // $X = [-(Ext-1)/2 : +(Ext-1)/2]$.
ThrHeightDiff // Height difference threshold in meters.
ThrSlope // Slope threshold in degrees.

Output:

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

Algorithm 4.3: Multi-directional slope dependent DTM generation (part 2).

```
1  Function DSM2DTM(DSM, Spacing, DSMsmooth, Dir, ExtendMetric,
   ThrHeightDiff, ThrSlope):
2      foreach ( scanline s ) do
                          // Loop over whole image in x and y direction.
3          DSMDiff = DSM(x,y) – DSM(x+ShiftX,y+ShiftY)      // Local
                                                  // height difference.
4          DSMsmoothDiff = DSMsmooth(x,y) –
            DSMsmooth(x+ShiftX,y+ShiftY)       // Local terrain slope.
5          Neigh = DSM(x+ExtX(Dir),y+ExtY(Dir))      // Neighborhood.
6          NeighCorr = Neigh + X*DSMsmoothDiff(x,y)         // Slope
                                              // corrected height values.
7          MinNeigh = min(NeighCorr)        // slope corrected minimal
                                                    // terrain value.
8          HeightDiff = DSM(x,y) – MinNeigh  // Difference to minimum.
9          if ( HeightDiff > ThrHeightDiff ) then
10             Label(x,y,s) = NonGround           // Pixel is non-ground.
11         else
12             Delta = DSMDiff – DSMsmoothDiff     // Slope corrected
                                                  // height difference.
13             SignDelta = -sign(Delta)          // Negated sign of Delta.
14             SlopeLocal = atan2(abs(Delta),Spacing))*180/pi   // Local
                                                          // slope.
15             Slope = SlopeLocal*SignDelta          // Corrected slope.
16             if ( Slope > ThrSlope ) then
17                 Label(x,y,s) = NonGround       // Pixel is non-ground.
18             else
19                 Label(x,y,s) = Label(x-ShiftX,y-ShiftY)     // Assign as
                                                        // last label.
20             end
21             if ( Slope < 0 ) then
22                 Label(x,y,s) = Ground                // Pixel is ground.
23             end
24         end
25     end
26     return Label                                  // Return classification.
```

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 are 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 and the resulting statistics are presented in Table 5.1. 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 (Astrium, 2012). 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 5.2. 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 3.1 and Figure 3.1). 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. For 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 5.3. 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, e.g., 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 (Jacobsen and Topan, 2015). 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 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 5.4. 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 5.1: 2D geo-location accuracy of Pléiades image data given in pixels. Shown are the initial (a-priori) mean and RMS residual errors.

Study Area	ID	GCPs	Mean initial [pxl]			RMS initial [pxl]		
			across	along	length	across	along	length
Innsbruck Stereo	1	30	-8.55	-5.91	10.43	8.61	5.96	10.47
	2	30	-8.55	-2.28	8.88	8.61	2.42	8.94
Innsbruck Triplet East	1	21	-9.61	-8.55	12.88	9.63	8.60	12.91
	2	20	-10.21	-9.49	13.96	10.24	9.54	14.00
	3	20	-10.14	-11.50	15.35	10.16	11.53	15.37
Innsbruck Triplet Center	1	16	-8.93	4.43	9.99	8.97	4.47	10.02
	2	16	-8.93	-0.18	8.97	8.95	0.91	8.99
	3	16	-9.10	4.18	10.03	9.13	4.22	10.06
Innsbruck Triplet West	1	22	-9.62	-8.56	12.92	9.67	8.59	12.94
	2	22	-9.23	-11.46	14.73	9.27	11.49	14.76
	3	22	-9.82	-9.53	13.70	9.85	9.55	13.72
Trento Triplet	1	21	11.19	12.59	16.90	11.22	12.70	16.95
	2	21	10.80	13.81	17.58	10.84	13.91	17.63
	3	21	8.61	15.71	17.97	8.65	15.81	18.02
Ljubljana Stereo 1	1	28	1.63	-0.05	1.83	1.81	0.65	1.93
	2	28	2.70	2.51	3.79	2.89	2.60	3.89
Ljubljana Stereo 2	3	30	1.41	-0.63	1.74	1.61	0.85	1.82
	4	30	1.42	1.72	2.44	1.81	1.88	2.61
Ljubljana Stereo 3	5	30	-3.07	11.66	12.07	3.13	11.67	12.08
	6	30	-2.93	12.14	12.50	2.99	12.16	12.52
Singapore Stereo	1	14	4.42	1.03	4.64	4.43	1.41	4.65
	2	16	4.28	-3.68	5.72	4.34	3.80	5.77

Next, the 3D geo-location accuracy using the initial and optimized sensor models is assessed. The RMS values of the 3D point residuals which were achieved for East, North, Height, and their 3D length are summarized in Table 5.5 and in Table 5.6. These values indicate the 3D mapping accu-

Table 5.2: 2D geo-location accuracy of Pléiades image data given in pixels. Shown are the RMS residual errors after parameter adjustment.

Study Area	ID	GCPs	RMS adjusted [pxl] across	along	length
Innsbruck Stereo	1	30	0.80	0.50	0.95
	2	30	0.78	0.61	1.00
Innsbruck Triplet East	1	21	0.57	0.80	0.99
	2	20	0.76	0.82	1.12
	3	20	0.53	0.79	0.96
Innsbruck Triplet Center	1	16	0.62	0.39	0.73
	2	16	0.44	0.65	0.78
	3	16	0.58	0.32	0.66
Innsbruck Triplet West	1	22	0.46	0.60	0.75
	2	22	0.66	0.62	0.91
	3	22	0.48	0.61	0.78
Trento Triplet	1	21	0.75	0.76	1.07
	2	21	0.67	0.71	0.98
	3	21	0.81	0.88	1.20
Ljubljana Stereo 1	1	28	0.55	0.60	0.82
	2	28	0.65	0.58	0.87
Ljubljana Stereo 2	3	30	0.54	0.52	0.75
	4	30	0.66	0.71	0.97
Ljubljana Stereo 3	5	30	0.63	0.60	0.87
	6	30	0.54	0.73	0.91
Singapore Stereo	1	14	0.27	0.88	0.92
	2	16	0.67	0.60	0.90

racy 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 δ, as an equivalent for the

Table 5.3: 2D geo-location accuracy analysis using different adjustment settings for the image 1 of the Innsbruck stereo test set. Given are mean and 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 5.4: 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

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 5.1 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.

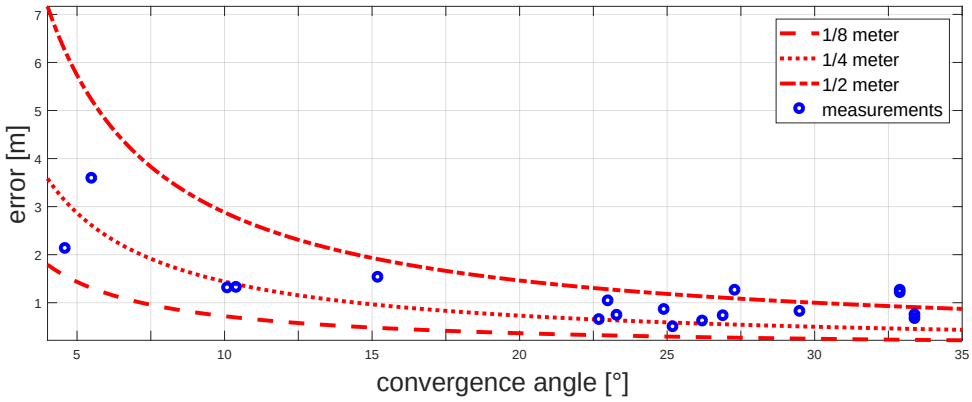


Figure 5.1: The 3D error is plotted versus the convergence angle, which are in the range from 4° to 35°, assuming an 2D location measurement error of 1/8, 1/4, and 1/2 m. On top of that the RMS values of the adjusted lengths (cf. Table 5.6) are shown in blue.

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, 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 Singpore 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 in height, if the reference GCPs are also in this accuracy range.

Table 5.5: 3D RMS residuals achieved from all initial sensor models given in meters. Values are given for East (E), North (N), Height (H), and their 3D Length (L).

Study Area	Model	Conv. Angle [°]	GCPs	RMS initial [m] E	N	H	L
Innsbruck Stereo	1-2	26.9	30	4.85	2.07	3.45	6.28
Innsbruck Triplet East	1-2	10.1	20	5.26	4.78	1.46	7.26
	2-3	23.3	20	5.44	5.10	2.79	7.96
	Triplet	33.4	20	5.33	5.14	2.31	7.76
Innsbruck Triplet Center	1-2	24.9	16	3.48	1.68	6.09	7.21
	2-3	4.6	16	3.85	2.07	4.38	6.19
	Triplet	29.5	16	3.41	1.64	5.83	6.95
Innsbruck Triplet West	1-2	10.4	22	5.03	4.72	2.17	7.24
	2-3	23.0	22	5.17	4.88	3.45	7.90
	Triplet	33.4	22	5.09	-4.96	2.97	7.70
Trento Triplet	1-2	5.5	21	4.75	8.14	6.40	11.39
	2-3	27.3	21	4.52	7.66	3.88	9.70
	1-3	32.9	21	4.60	7.60	4.16	9.81
	Triplet	32.9	21	4.70	7.66	4.07	9.87
Ljubljana Stereo 1	1-2	25.2	28	0.99	0.71	2.72	2.98
Ljubljana Stereo 2	3-4	22.7	30	1.16	0.58	2.90	3.17
Ljubljana Stereo 3	5-6	26.2	30	1.54	6.58	1.02	6.83
Singapore Stereo	1-2	15.2	13	1.20	1.82	9.19	9.44

Analogue 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

Table 5.6: 3D RMS residuals achieved from all adjusted sensor models given in meters. Values are given for East (E), North (N), Height (H), and their 3D Length (L).

Study Area	Model	Conv. Angle [°]	GCPs	RMS adjusted [m] E	N	H	L
Innsbruck Stereo	1-2	26.9	30	0.38	0.23	0.59	0.73
Innsbruck Triplet East	1-2	10.1	20	0.33	0.46	1.18	1.31
	2-3	23.3	20	0.29	0.39	0.56	0.74
	Triplet	33.4	20	0.28	0.39	0.47	0.67
Innsbruck Triplet Center	1-2	24.9	16	0.36	0.19	0.76	0.86
	2-3	4.6	16	0.59	0.36	2.01	2.13
	Triplet	29.5	16	0.33	0.17	0.73	0.82
Innsbruck Triplet West	1-2	10.4	22	0.19	0.33	1.26	1.32
	2-3	23.0	22	0.26	0.27	0.98	1.04
	Triplet	33.4	22	0.21	0.25	0.67	0.75
Trento Triplet	1-2	5.5	21	0.94	0.88	3.35	3.59
	2-3	27.3	21	0.36	0.34	1.15	1.26
	1-3	32.9	21	0.42	0.31	1.15	1.26
	Triplet	32.9	21	0.38	0.31	1.11	1.21
Ljubljana Stereo 1	1-2	25.2	28	0.29	0.30	0.28	0.50
Ljubljana Stereo 2	3-4	22.7	30	0.30	0.32	0.48	0.65
Ljubljana Stereo 3	5-6	26.2	30	0.31	0.37	0.40	0.62
Singapore Stereo	1-2	15.2	13	0.31	0.29	1.47	1.53

test site, tests with a different number of GCPs and ICPs as well as a 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 5.7. 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, e.g., 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 5.7: 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 5.8 and in Table 5.9. 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 5.4). 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 5.2 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, namely the pairs 1-3, 2-4, 3-5, and 4-6. When compared to the acquisition disposition depicted in Figure 3.3 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.

Table 5.8: 3D RMS residuals using GCPs and ICPs for the Ljubljana test set given in meters (part 1). 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).

Name	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

Table 5.9: 3D RMS residuals using GCPs and ICPs for the Ljubljana test set given in meters (part 2). 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).

Name	Type	RMS initial [m]				RMS adjusted [m]			
		E	N	H	L	E	N	H	L
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

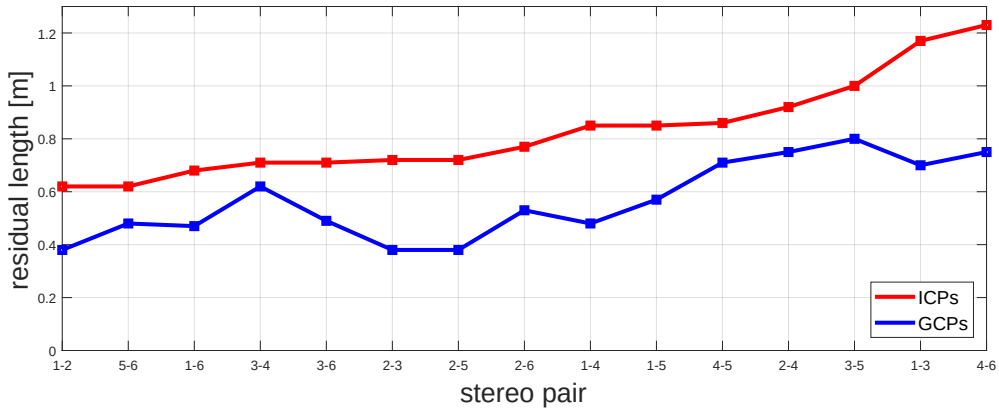


Figure 5.2: The 3D length discrepancies for the Ljubljana set sorted from lowest to highest ICP error. ICP values are given in red and corresponding GCP values in blue.

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. (Piermattei, Marty, Karel, et al., 2018)). 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 5.10 shows the statistics of the deviation of tie-points orthogonal to the according 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. (Piermattei, Marty, Karel, et al., 2018)). Each TP between two images yield four additional equations in our non-linear equation system (cf. Eq. (4.18) and Algorithm 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)* (Lowe, 2004), *speeded up robust features (SURF)* (Bay, Tuytelaars, and Van Gool, 2006), or similar techniques (Tuytelaars and Mikolajczyk, 2008). In this case, a block adjustment can be performed for multiple images and thus for multiple RPC sensor models simultaneously.

Table 5.10: 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 evaluation is limited to the three prime test sites Innsbruck, Trento, and Ljubljana. For image matching the parameters as given in Table 5.11 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 is 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 5.11: 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.

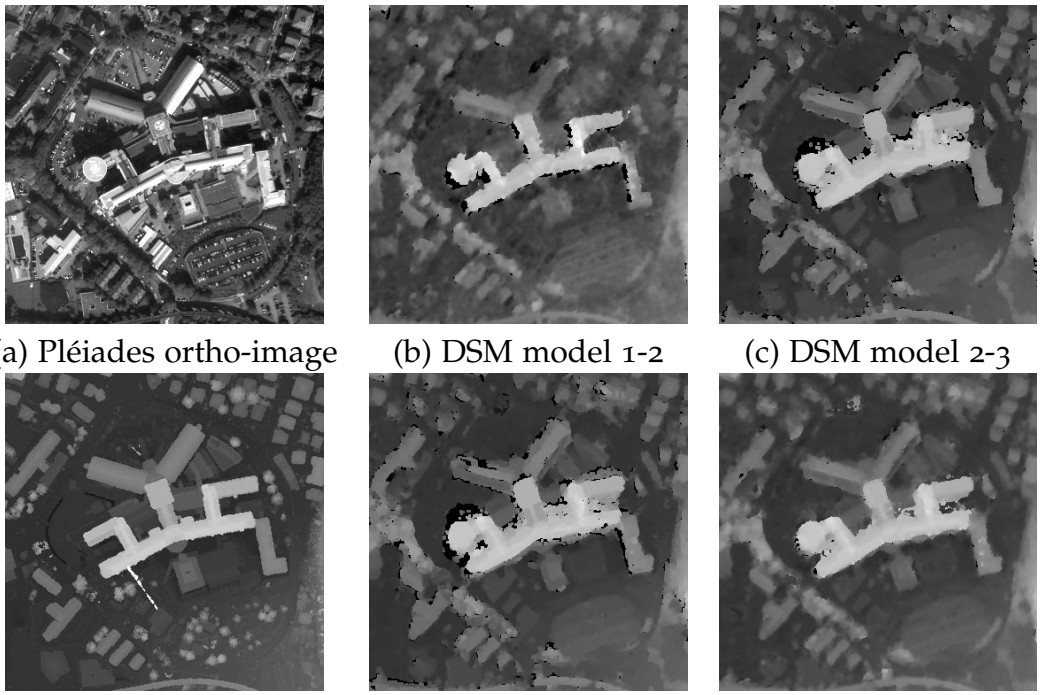
Parameter	**Standard version**	**Accurate version**
Matcher	SGM	SGM
Cost function	Census 9×9 pixel	**Census 9×9 pixel and SAD 3×3 pixel**
Prediction	SRTM DEM	SRTM DEM
Penalties	$P_1 = 0.4$, $P_2' = 1.5$	$P_1 = 0.4$, $P_2' = 1.5$
Pyramid levels	4	4
Search size	41 / 15 / 15	**41 / 31 / 31**
Cost threshold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

Figure 5.3 shows a subset of the Trento test site, which covers a hospital and its surrounding area of 360×360 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 5.12. The STDs 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 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.

Figure 5.4 shows a subset of the Innsbruck test site, which covers an urban region of 1950×1000 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



(a) Pléiades ortho-image (b) DSM model 1-2 (c) DSM model 2-3

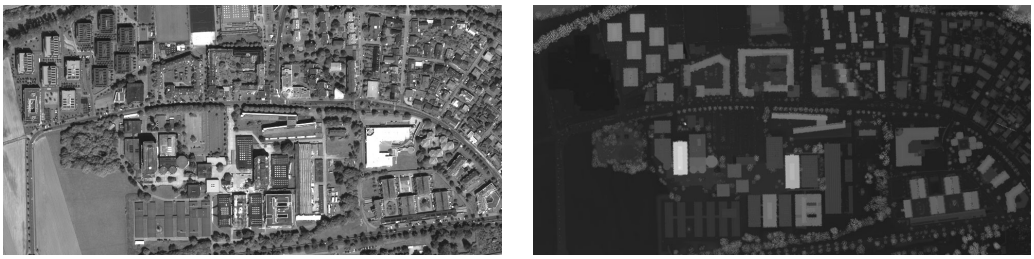
(d) LiDAR DSM (e) DSM model 1-3 (f) Triplet DSM

Figure 5.3: Detail view of a hospital at the Trento test site.

large buildings are decently reconstructed while they are missing in the 2D matching based DSM.

Table 5.12: Statistics of height differences between LiDAR and Pléiades DSMs on non-forest regions.

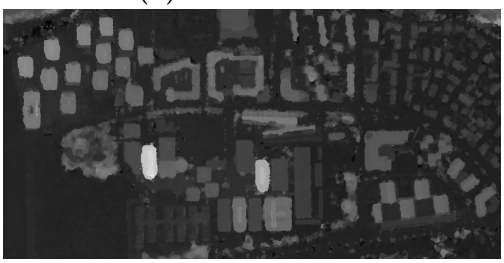
Study Area	**Model**	**Mean [m]**	**STD [m]**	**Area [km^2]**
Trento Triplet	1-2	-0.62	4.53	10.8
	2-3	-0.55	1.97	10.8
	1-3	-0.53	2.25	10.8
	Triplet	-0.56	1.99	10.8
Innsbruck Stereo	1-2	0.25	1.68	26.5



(a) Pléiades ortho-image

(b) LiDAR DSM

(c) DSM based on 2D matching

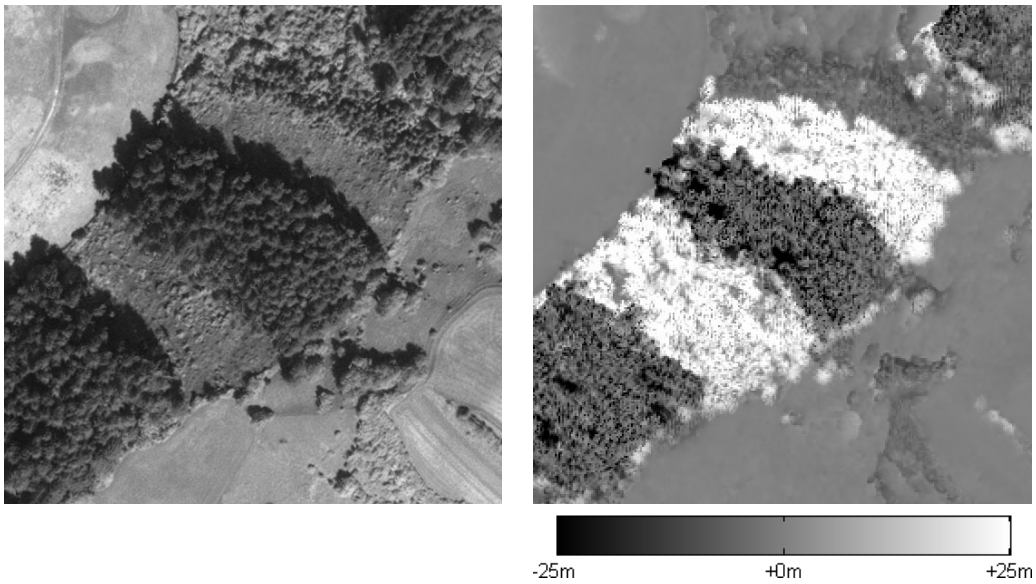


(d) DSM based on the proposed workflow

Figure 5.4: Detail view of an urban area at the Innsbruck stereo test site.

For a partly forested region at test site Innsbruck with 550×550 m^2 Figure 5.5 shows a Pléiades ortho-image as well as a 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.

Analogue 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 5.12 and show a 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 5.6). 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.



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

Figure 5.5: Detail view of a partly forested area at the Innsbruck stereo test site.

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 3.2. 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 5.13 lists a multitude of statistical values, in particular minimal, maximal, mean, STD, MED, NMAD, mean absolute error (MAE), percentage of nodata values (i.e., $100\% - \text{completeness}$), and the convergence angles. Figure 5.6 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 (Perko, Schardt, et al., 2019)).

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 3.3) 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 values is not the best idea for our difference models as they are non-Gaussian distributed.

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 5.7. 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, e.g., 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, i.e., 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 5.8 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.

Table 5.13: Accuracy analysis of the resulting DSMs in comparison to LiDAR reference data. 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 [%]	δ [°]
1-2	-895.1	769.2	-0.36	**4.33**	0.06	0.97	2.05	14.34	25.2
1-3	-895.5	627.5	-1.06	4.47	-0.49	1.46	2.33	**6.60**	12.2
1-4	-893.8	627.9	-0.39	4.51	0.07	0.94	**2.03**	18.77	27.5
1-5	-891.4	630.4	-0.53	4.48	**0.03**	1.20	2.22	18.07	24.0
1-6	-894.1	**626.8**	-0.47	5.48	-0.24	**0.86**	2.25	27.55	33.9
2-3	-895.7	628.8	-0.26	4.47	0.07	1.04	2.15	16.76	26.4
2-4	-896.2	632.4	-0.40	4.47	0.08	1.46	2.26	7.76	12.5
2-5	-893.5	628.4	0.09	5.39	0.28	0.87	2.36	28.00	37.6
2-6	-382.9	627.9	-0.17	4.90	**0.03**	1.14	2.38	22.77	25.0
3-4	-895.0	627.5	**-0.06**	4.49	0.22	1.01	2.13	14.82	22.7
3-5	**-381.5**	629.0	-0.09	4.56	0.43	1.58	2.44	10.72	13.0
3-6	-894.7	627.4	-0.61	4.53	-0.24	1.04	2.13	19.06	24.7
4-5	-894.4	628.1	0.19	4.90	0.37	0.93	2.30	25.05	30.3
4-6	-897.3	630.2	-0.66	4.75	-0.23	1.69	2.52	12.28	12.6
5-6	-893.1	766.0	-0.48	4.61	-0.08	0.92	2.09	21.22	26.2
1-2, 3-4	-895.1	628.3	-0.20	4.41	0.15	0.91	2.07	12.21	many
1-2, 3-4, 5-6	-894.6	627.6	-0.29	4.37	0.08	0.84	2.01	12.85	many
$\delta \in [0, 20]$	-896.4	626.7	-0.51	4.07	**-0.01**	1.14	1.98	11.07	many
$\delta \in [20, 30]$	-894.2	627.9	-0.37	4.56	0.05	0.93	2.16	7.30	many
$\delta \in [30, 40]$	**-894.0**	**575.6**	**-0.05**	**2.68**	0.16	**0.41**	**1.00**	55.19	many
all	-894.6	765.4	-0.46	4.74	0.05	1.09	2.32	**1.12**	all

Just for visual purposes a subset of the Tian Shui test site is depicted in Figure 5.9 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.

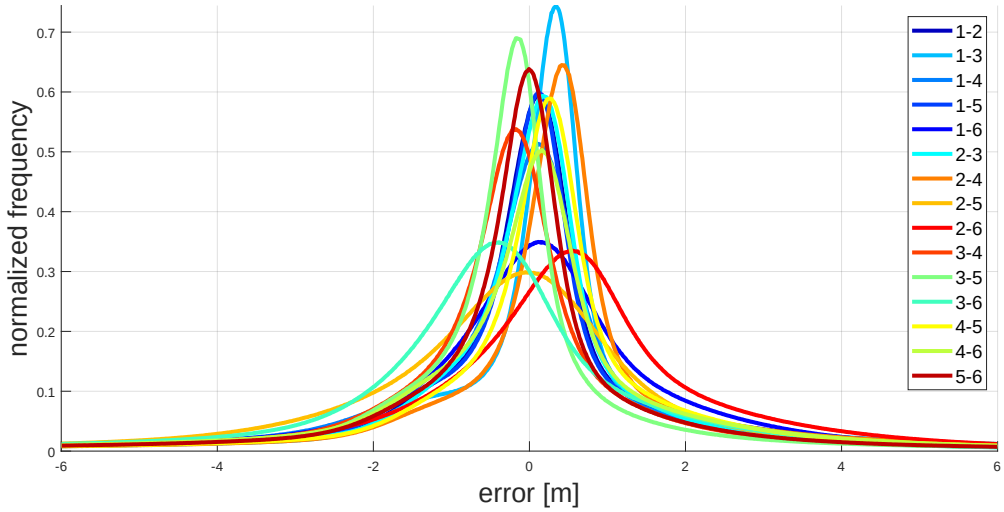


Figure 5.6: Distribution of differences from LiDAR and Pléiades DSMs for all 15 stereo pairs of the Ljubljana test site. The error range of -6 m to $+6$ m is shown.

5.3 Digital Terrain Model Generation

Tests are preformed 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 5.14 have been used.

Table 5.14: Parameters for terrain model extraction as used in all tests.

Parameter	Value
Filter extent	91 meters
Height threshold	3 meters
Slope threshold	30 degrees

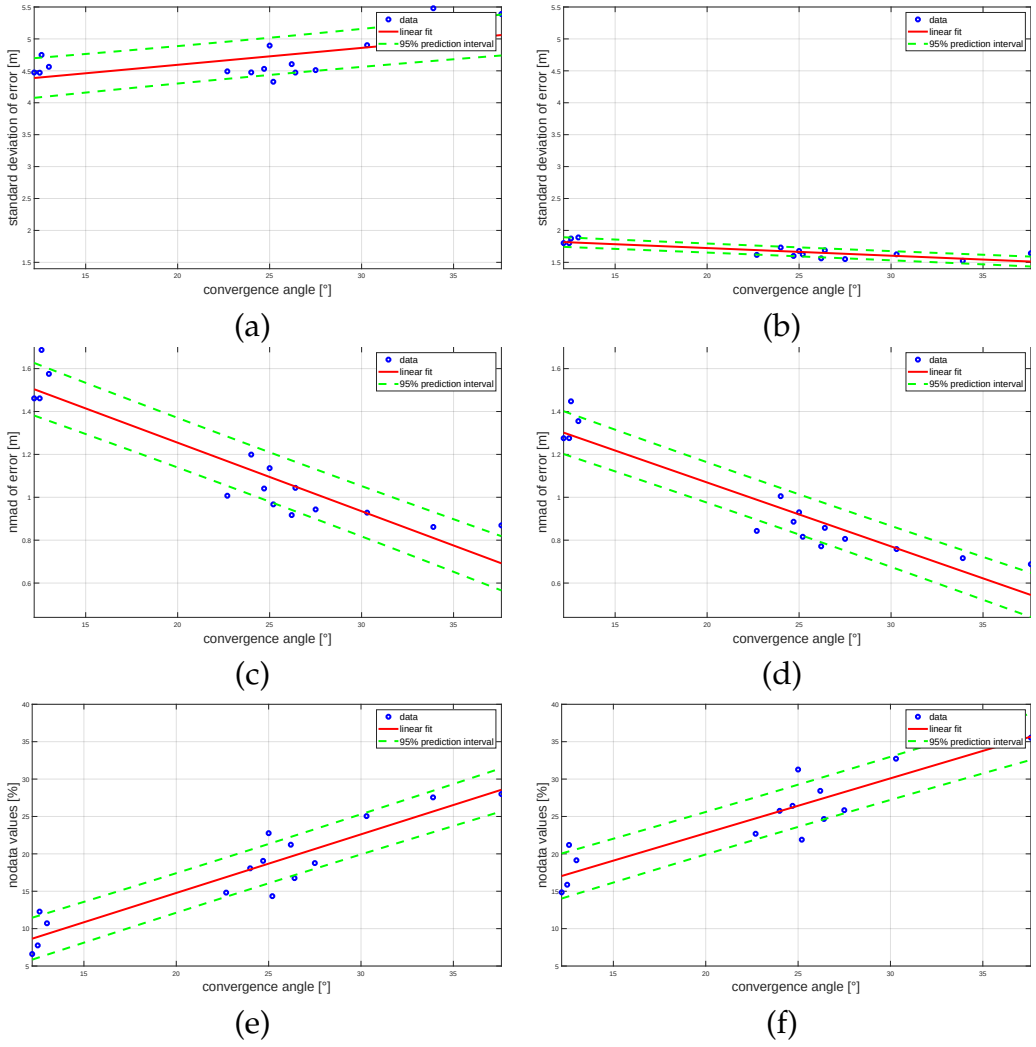


Figure 5.7: Plots showing (a,b) STD, (c,d) NMAD, and (e,f) percentage of nodata values plotted versus the convergence angle for the Ljubljana test site. (a,c,e) are based on the complete error distribution, while (b,d,f) is based on a clipped distribution, where all gross outliers with an absolute value larger than 6 m are removed.

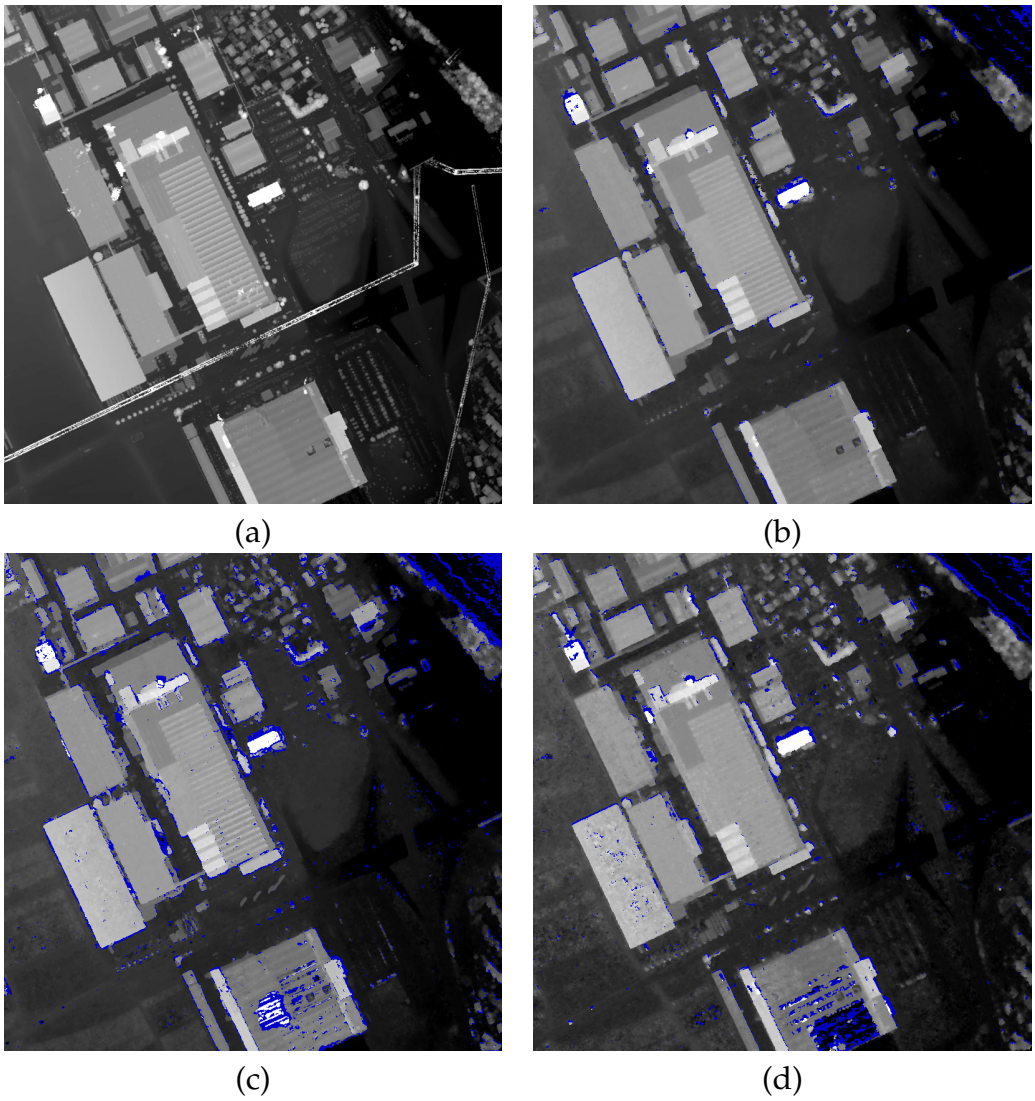
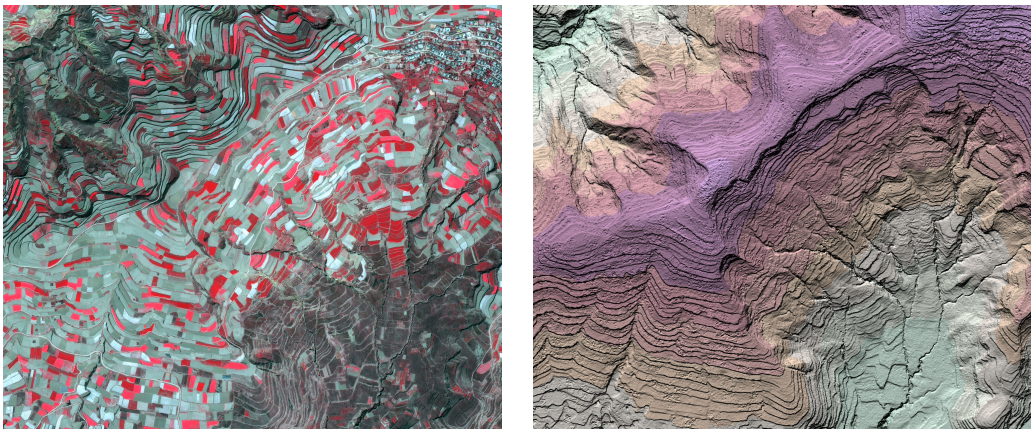


Figure 5.8: Selected DSM subsets with 800×800 m^2 and 1 m GSD of the Ljubljana test site. Height ranges from 410 m to 500 m. Shown are (a) LiDAR reference, (b) Pléiades DSM based on all stereo pairs with convergence angles in the range of 20° to 30°, (c) Pléiades DSM based on stereo pair 1-2, and (d) Pléiades DSM based on stereo pair 1-3. Dark blue pixel indicate non-reconstructed regions.

5.3.1 Qualitative Evaluation

For visual comparison the results of several ground pixel filtering methods are given in Figure 5.10 for an area of 721×361 m^2, showing a residential



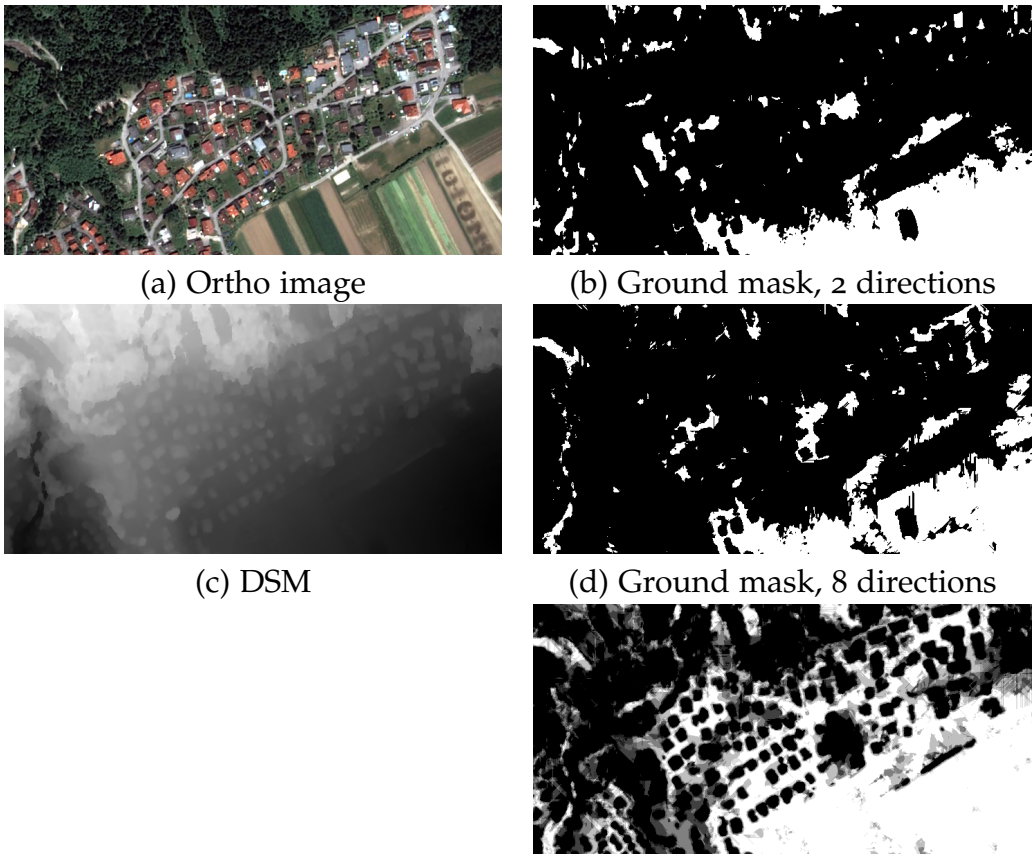
(a) (b)

Figure 5.9: Ortho CIR image (a) and relief shaded DSM (b) are visualized for a subset of the Tian Shui test site covering 3.4 km^2.

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 (Meng, L. Wang, et al., 2009)), (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 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 5.11 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 5.12 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. Due to the large temporal gap between LiDAR and Pléiades data acquisitions only selected areas are analyzed, which are not affected by temporal changes



(a) Ortho image

(b) Ground mask, 2 directions

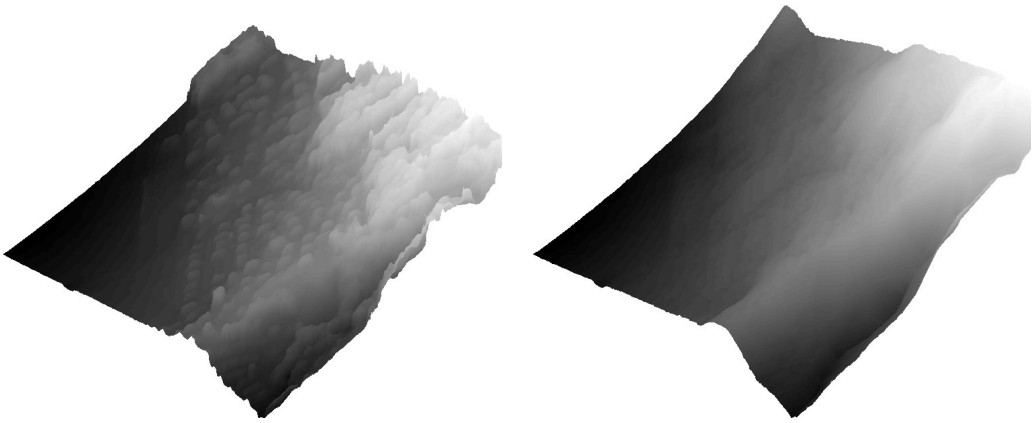
(c) DSM

(d) Ground mask, 8 directions

(e) Ground mask, 8 directions, slope dependent

Figure 5.10: Subset of (a) true-ortho image, (c) input DSM, and (b,d,e) together with ground masks using different methods.

due to construction, vegetation growth or cloud cover. Mean values as well as standard deviations of the height differences are summarized in Table 5.15 for an area of 25.4 km^2. 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 (Stumpf et al., 2014; Perko, Raggam, Gutjahr, et al., 2014; Berthier et al., 2014). Consequently, the differences of LiDAR and Pléiades based DTMs will have the same bias, which however is not a result of the presented DTM generation algorithm.



(a) (b)

Figure 5.11: 3D view of (a) the DSM as shown in Figure 5.10 and (b) the resulting DTM after applying the proposed filtering method.

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 5.15. The mean bias between both DTMs indeed is very similar to the one achieved for the DSMs. Due to the fact, that 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.

Table 5.15: 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 5.13 shows DSMs and DTMs as generated from LiDAR and Pléiades data for two small areas of 100 × 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 gives indication, that small structures, like

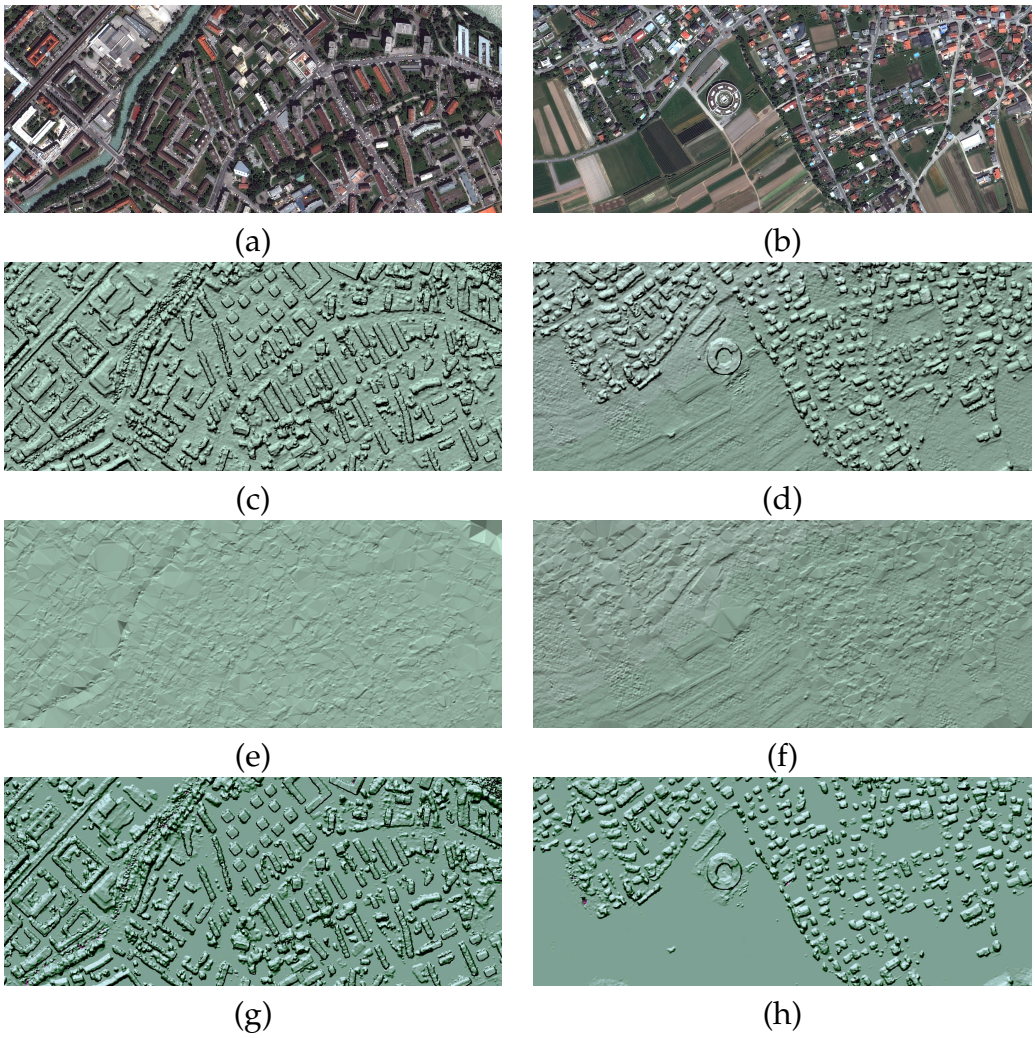


Figure 5.12: Urban subset (left) and suburban dataset (right). Shown are (a,b) ortho image, (c,d) extracted DSM, (e,f) derived DTM, and (g,h) according nDSM in relief view.

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.

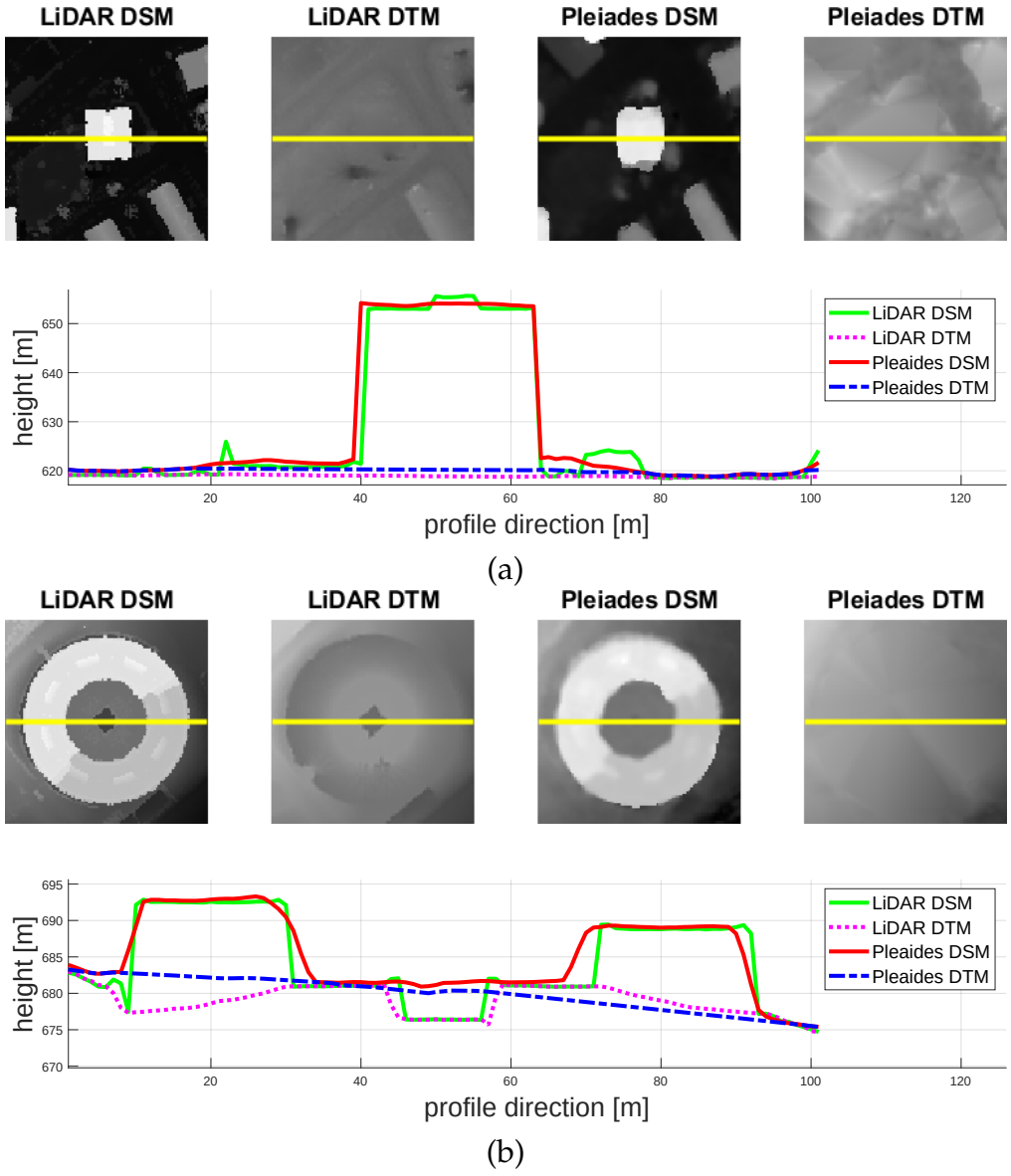


Figure 5.13: Profile comparisons of DSMs and DTMs from LiDAR and Pléiades for two small scenes, representing (a) a skyscraper and (b) a round building.

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 minimal 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 post-processing small patches in the ground mask are removed by applying a peak filter (cf. Figure 5.10), since, e.g., 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, e.g., applying a Gaussian, a median, or a bilateral filter (Tomasi and Manduchi, 1998).

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 thesis 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 (Gutjahr et al., 2014) was 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 a downsampling within the fusion step.

13. Novel multi-directional slope dependent DTM generation method, which extends the approach in (Meng, L. Wang, et al., 2009).

Additionally, all steps of the workflow were evaluated by in-depth accuracy assessments based on a multitude of satellite and according meta data. The assessment itself represents a main contribution of this thesis. Since the presented methods were implemented within a commercially available software interested audience can directly build upon the findings. Note, that the presented workflow was already successfully applied by other research groups, for instance, in (Deutscher et al., 2013; Persson, 2016; Persson and Perko, 2016; Himmelreich, 2017; Leopold et al., 2017). Especially, the DTM generation was applied in (Mousa, Helmholz, and Belton, 2017; Luethje, Tiede, and Eisank, 2017; Auer, Schmitt, and Reinartz, 2017; Misra and Takeuchi, 2017; Auer, Reinartz, and Schmitt, 2018; Panagiotakis et al., 2018; Mousa, Helmholz, Belton, and Bulatov, 2019).

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, a 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 are 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. 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 exists only stereo pairs with appropriate convergence angle should be used. The comparisons of LiDAR to Pléiades based DSMs showed highly accurate reconstructions, with a 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 forest 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 similar quality than other commercial software packages.

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 a RMF could be solved in a heuristic manner. 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, e.g., by a variational approach. In addition, the multi-spectral information, i.e., our ortho mosaic may serve as a 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. Beside 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. Last but not least, the main scientific achievement of this thesis are the three accepted publications (Perko, Raggam, Schardt, et al., 2018; Perko, Hirschmugl, Deutscher, et al., 2019; Perko, Schardt, et al., 2019) and the publication currently under review (Perko, Raggam, and Roth,

2019). They show that the topic of this thesis is of interest for the scientific community and demonstrate the scientific level of the proposed end-to-end mapping workflow.

Bibliography

Abduelmola, Abdunaser Ebrahime Aboajela (2016). "High Resolution Satellite Image Analysis and Rapid 3D Model Extraction for Urban Change Detection." PhD thesis. University of Porto (cit. on p. 1).

Aguilar, Manuel Ángel, Maria del Mar Saldaña, and Fernando José Aguilar (2014). "Generation and quality assessment of stereo-extracted DSM from GeoEye-1 and WorldView-2 imagery." In: *IEEE Transactions on Geoscience and Remote Sensing* 52.2, pp. 1259–1271 (cit. on pp. 12, 36).

Ahlberg, Carl et al. (2019). "Unbounded Sparse Census Transform Using Genetic Algorithm." In: *IEEE Winter Conference on Applications of Computer Vision*, pp. 1616–1625 (cit. on p. 15).

Airbus Defence and Space (2018). *Pleiades Neo Leaflet*. Tech. rep. https://www.intelligence-airbusds.com/files/pmedia/public/r51130_9_leaflet-pleiadesneov2.pdf (cit. on p. 6).

Arefi, Hossein et al. (2011). "Iterative approach for efficient digital terrain model production from CARTOSAT-1 stereo images." In: *Journal of Applied Remote Sensing* 5.1, pp. 1–19 (cit. on p. 18).

Åstrand, Pär Johan et al. (2012). "The potential of WorldView-2 for orthoimage production within the "Control with Remote Sensing Programme" of the European Commission." In: *International Journal of Applied Earth Observation and Geoinformation* 19, pp. 335–347 (cit. on pp. 12, 36).

Astrium (2012). *Pleiades Technical Documents: Pleiades User Guide V 2.0*. Tech. rep. https://www.intelligence-airbusds.com/en/4572-pleiades-technical-documents (cit. on pp. 4–6, 59).

Auer, Stefan, Peter Reinartz, and Michael Schmitt (2018). "Object-related alignment of heterogeneous image data in remote sensing." In: *International Conference on Information Fusion*, pp. 1607–1614 (cit. on p. 90).

Auer, Stefan, Michael Schmitt, and Peter Reinartz (2017). "Automatic alignment of high resolution optical and SAR images for urban areas." In: *IEEE International Geoscience and Remote Sensing Symposium*, pp. 5466–5469 (cit. on p. 90).

Axelsson, Peter (2000). "DEM generation from laser scanner data using adaptive TIN models." In: *International Archives of Photogrammetry and Remote Sensing* 33.4, pp. 110–117 (cit. on p. 17).

Bagnardi, Marco, Pablo J González, and Andrew Hooper (2016). "High-resolution digital elevation model from tri-stereo Pleiades-1 satellite imagery for lava flow volume estimates at Fogo Volcano." In: *Geophysical Research Letters* 43.12, pp. 6267–6275 (cit. on p. 1).

Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool (2006). "SURF: Speeded up robust features." In: *European Conference on Computer Vision*, pp. 404–417 (cit. on p. 70).

Belart, Joaquín MC et al. (2019). "The geodetic mass balance of Eyjafjallajökull ice cap for 1945–2014: processing guidelines and relation to climate." In: *Journal of Glaciology*, pp. 1–15 (cit. on p. 1).

Belgiu, Mariana, Lucian Drăguţ, and Josef Strobl (2014). "Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using WorldView-2 imagery." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 87, pp. 205–215 (cit. on p. 1).

Bernard, Marc et al. (2012). "3D capabilities of Pleiades satellite." In: *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 39, pp. 553–557 (cit. on p. 1).

Berthier, Etienne et al. (2014). "Glacier topography and elevation changes from Pléiades very high resolution stereo images." In: *The Cryosphere* 8.5, pp. 4849–4883 (cit. on pp. 2, 83).

Bittner, Ksenia et al. (2018). "DSM-to-LoD2: Spaceborne Stereo Digital Surface Model Refinement." In: *Remote Sensing* 10.12, p. 1926 (cit. on p. 1).

Bleyer, Michael and Sylvie Chambon (2010). "Does color really help in dense stereo matching." In: *International Symposium 3D Data Processing, Visualization and Transmission*, pp. 1–8 (cit. on p. 40).

Bosch, Marc, Zachary Kurtz, et al. (2016). "A multiple view stereo benchmark for satellite imagery." In: *Applied Imagery Pattern Recognition Workshop*, pp. 1–9 (cit. on p. 2).

Bosch, Marc, A Leichtman, et al. (2017). "Metric Evaluation Pipeline for 3D Modeling of Urban Scenes." In: *International Archives of the Photogramme-*

try, Remote Sensing & Spatial Information Sciences 42, pp. 239–246 (cit. on p. 1).

Bredies, Kristian, Karl Kunisch, and Thomas Pock (2010). "Total generalized variation." In: *SIAM Journal on Imaging Sciences* 3.3, pp. 492–526 (cit. on p. 55).

Campbell, James B and Randolph H Wynne (2011). *Introduction to Remote Sensing*. Guilford Press (cit. on p. 19).

Capaldo, Paola et al. (2012). "DSM generation from high resolution imagery: Applications with WorldView-1 and Geoeye-1." In: *Italian Journal of Remote Sensing* 44.1, pp. 41–53 (cit. on p. 1).

D'Angelo, Pablo and Georg Kuschk (2012). "Dense multi-view stereo from satellite imagery." In: *IEEE International Geoscience and Remote Sensing Symposium*, pp. 6944–6947 (cit. on p. 2).

Davies, E Roy (2012). *Computer and Machine Vision: Theory, Algorithms, Practicalities*. Academic Press (cit. on p. 16).

De Franchis, Carlo et al. (2014a). "An automatic and modular stereo pipeline for pushbroom images." In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 49–56 (cit. on p. 13).

De Franchis, Carlo et al. (2014b). "Automatic sensor orientation refinement of Pléiades stereo images." In: *IEEE Geoscience and Remote Sensing Symposium*, pp. 1639–1642 (cit. on p. 13).

Deutscher, Janik et al. (2013). "Mapping tropical rainforest canopy disturbances in 3D by COSMO-SkyMed Spotlight InSAR-stereo data to detect areas of forest degradation." In: *Remote sensing* 5.2, pp. 648–663 (cit. on p. 90).

Dial, Gene and Jacek Grodecki (2002). "Block adjustment with rational polynomial camera models." In: *American Society for Photogrammetry and Remote Sensing Annual Conference*, pp. 22–26 (cit. on pp. 11, 12, 31, 33, 36).

Drory, Amnon et al. (2014). "Semi-global matching: a principled derivation in terms of message passing." In: *German Conference on Pattern Recognition*, pp. 43–53 (cit. on pp. 14, 40).

Duan, Liuyun and Florent Lafarge (2016). "Towards large-scale city reconstruction from satellites." In: *European Conference on Computer Vision*, pp. 89–104 (cit. on p. 1).

Đurić, Dragana et al. (2017). "Using multiresolution and multitemporal satellite data for post-disaster landslide inventory in the Republic of Serbia." In: *Landslides* 14.4, pp. 1467–1482 (cit. on p. 1).

Eineder, Michael et al. (2010). "Imaging geodesy—Toward centimeter-level ranging accuracy with TerraSAR-X." In: *IEEE Transactions on Geoscience and Remote Sensing* 49.2, pp. 661–671 (cit. on p. 12).

Facciolo, Gabriele, Carlo De Franchis, and Enric Meinhardt-Llopis (2017). "Automatic 3D reconstruction from multi-date satellite images." In: *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 57–66 (cit. on p. 2).

Facciolo, Gabriele, Carlo De Franchis, and Enric Meinhardt (2015). "MGM: A significantly more global matching for stereovision." In: *British Machine Vision Conference*, pp. 1–12 (cit. on p. 14).

Farr, Tom G et al. (2007). "The shuttle radar topography mission." In: *Reviews of Geophysics* 45.2 (cit. on p. 43).

Fife, Wade S and James K Archibald (2013). "Improved census transforms for resource-optimized stereo vision." In: *IEEE Transactions on Circuits and Systems for Video Technology* 23.1, pp. 60–73 (cit. on p. 15).

Förstner, Wolfgang and Bernhard P Wrobel (2016). *Photogrammetric Computer Vision*. Springer (cit. on p. 19).

Fourest, Sébastien et al. (2012). "Star-based methods for Pleiades HR commissioning." In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 39, pp. 513–518 (cit. on p. 4).

Fox, John and Sanford Weisberg (2018). *An R companion to applied regression*. Sage Publications (cit. on p. 52).

Fraser, Clive S and Harry B Hanley (2003). "Bias compensation in rational functions for IKONOS satellite imagery." In: *Photogrammetric Engineering & Remote Sensing* 69.1, pp. 53–57 (cit. on p. 12).

Fusiello, Andrea and Luca Irsara (2008). "Quasi-euclidean uncalibrated epipolar rectification." In: *International Conference on Pattern Recognition*, pp. 1–4 (cit. on p. 13).

Fusiello, Andrea, Emanuele Trucco, and Alessandro Verri (2000). "A compact algorithm for rectification of stereo pairs." In: *Machine Vision and Applications* 12.1, pp. 16–22 (cit. on p. 13).

Gleyzes, Alain, Lionel Perret, and Eric Cazala-Houcade (2013). "Pleiades system fully operational in orbit." In: *EARSeL Symposium*. Vol. 33, pp. 445–460 (cit. on p. 4).

Gong, Kang and Dieter Fritsch (2019). "DSM Generation from High Resolution Multi-View Stereo Satellite Imagery." In: *Photogrammetric Engineering & Remote Sensing* 85.5, pp. 379–387 (cit. on p. 2).

Greenwalt, Clyde R and Melvin E Shultz (1962). *Principles of error theory and cartographic applications*. Tech. rep. Aeronautical Chart and Information Center St Louis MO (cit. on p. 6).

Grodecki, Jacek and Gene Dial (2003). "Block adjustment of high-resolution satellite images described by rational polynomials." In: *Photogrammetric Engineering & Remote Sensing* 69.1, pp. 59–68 (cit. on pp. 12, 31, 33).

Gutjahr, Karlheinz et al. (2014). "The epipolarity constraint in stereo-radargrammetric DEM generation." In: *IEEE Transactions on Geoscience and Remote Sensing* 52.8, pp. 5014–5022 (cit. on pp. 1, 9, 13, 37, 89).

Haller, István and Sergiu Nedevschi (2012). "Design of Interpolation Functions for Subpixel-Accuracy Stereo-Vision Systems." In: *IEEE Transactions on Image Processing* 21.2, pp. 889–898 (cit. on p. 44).

Hartley, Richard and Andrew Zisserman (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd ed (cit. on p. 5).

Hermann, Simon and Reinhard Klette (2009). "The naked truth about cost functions for stereo matching." In: *Univ. Auckland, Auckland, New Zealand, MI-tech-TR* 33 (cit. on p. 40).

Himmelreich, Laura Christin (2017). "DHM Ableitungen aus Pléiades Tri-Stereo Satellitenbildern im Hochgebirge. Digitale Höhenmodelle Verschiedener Softwareprodukte im Vergleich zu ALS Daten." MA thesis. University of Innsbruck, Innsbruck, Austria (cit. on p. 90).

Himmelreich, Laura Christin, Martin Ladner, and Armin Heller (2017). "Pléiades Tri-Stereo-Bilder im Hochgebirge - Eine Parameterstudie mit PCI Geomatics." In: *AGIT Journal* 3, pp. 153–162 (cit. on p. 1).

Hirschmugl, Manuela et al. (2005). "Pansharpening - Methoden für digitale, sehr hoch auflösende Fernerkundungsdaten." In: *Symposium und Fachmesse für Angewandte GeoInformatik*, pp. 270–276 (cit. on p. 4).

Hirschmüller, Heiko (2008). "Stereo processing by semiglobal matching and mutual information." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.2, pp. 328–341 (cit. on pp. 13, 14, 16, 40–42, 45, 53).

Hirschmüller, Heiko and Daniel Scharstein (2007). "Evaluation of cost functions for stereo matching." In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (cit. on pp. 14, 40).

Höhle, Joachim and Michael Höhle (2009). "Accuracy assessment of digital elevation models by means of robust statistical methods." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 64.4, pp. 398–406 (cit. on p. 47).

Isenburg, Martin (2013). "LASzip: lossless compression of LiDAR data." In: *Photogrammetric Engineering and Remote Sensing* 79.2, pp. 209–217 (cit. on p. 45).

Jacobsen, Karsten (2011). "Characteristics of very high resolution optical satellites for topographic mapping." In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. 38, pp. 137–142 (cit. on p. 4).

Jacobsen, Karsten and Hüseyin Topan (2015). "DEM generation with short base length Pleiades triplet." In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 40, pp. 81–86 (cit. on pp. 12, 22, 33, 47, 60).

Keys, Robert (1981). "Cubic convolution interpolation for digital image processing." In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 29.6, pp. 1153–1160 (cit. on p. 37).

Kim, Taejung (2000). "A study on the epipolarity of linear pushbroom images." In: *Photogrammetric Engineering & Remote Sensing* 66.8, pp. 961–966 (cit. on p. 13).

Klopschitz, Manfred et al. (2017). "Projected Texture Fusion." In: *International Symposium on Image and Signal Processing and Analysis*. Vol. 10, pp. 109–114 (cit. on pp. 14, 40).

Knöbelreiter, Patrick, Christoph Vogel, and Thomas Pock (2018). "Self-Supervised Learning for Stereo Reconstruction on Aerial Images." In: *IEEE International Geoscience and Remote Sensing Symposium*, pp. 4379–4382 (cit. on p. 15).

Kolmogorov, Vladimir and Ramin Zabih (2001). "Computing visual correspondence with occlusions using graph cuts." In: *IEEE International Conference on Computer Vision*. Vol. 2, pp. 508–515 (cit. on p. 14).

Kothencz, Gyula et al. (2018). "Urban vegetation extraction from VHR (tri-) stereo imagery - A comparative study in two central European cities." In: *European Journal of Remote Sensing* 51.1, pp. 285–300 (cit. on p. 1).

Kraus, Karl (2011). *Photogrammetry: Geometry from Images and Laser Scans*. Walter de Gruyter (cit. on p. 19).

Kraus, Karl and Norbert Pfeifer (2001). "Advanced DTM generation from LIDAR data." In: *International Archives Of Photogrammetry Remote Sensing and Spatial Information Sciences* 34.3/W4, pp. 23–30 (cit. on p. 17).

Krauß, Thomas, Hossein Arefi, and Peter Reinartz (2011). "Evaluation of selected methods for extracting digital terrain models from satellite born digital surface models in urban areas." In: *International Conference on Sensors and Models in Photogrammetry and Remote Sensing*. Vol. 5, pp. 1–7 (cit. on p. 18).

Krauß, Thomas, Pablo D'Angelo, and Lorenz Wendt (2018). "Cross-track satellite stereo for 3D modelling of urban areas." In: *European Journal of Remote Sensing*, pp. 1–10 (cit. on p. 2).

Krauß, Thomas and Peter Reinartz (2010). "Urban object detection using a fusion approach of dense urban digital surface models and VHR optical satellite stereo data." In: *ISPRS Istanbul Workshop*, pp. 1–6 (cit. on p. 18).

Ladner, Martin, Armin Heller, and Ekkehart Grillmayer (2017). "GNSS und Pléiades-Bilder in der Hochgebirgskartographie." In: *AGIT Journal* 3, pp. 13–23 (cit. on p. 1).

Leberl, Franz et al. (2003). "The UltraCam large format aerial digital camera system." In: *American Society For Photogrammetry & Remote Sensing*, pp. 5–9 (cit. on p. 18).

Leopold, Philip et al. (2017). "Comparing landslide mapping from DTM satellite derived data and field based studies of Loess sediments in Western China." In: *World Landslide Forum 4 - Advancing Culture of Living with Landslides* 2, pp. 87–95 (cit. on pp. 1, 28, 90).

Leys, Christophe et al. (2013). "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median." In: *Journal of Experimental Social Psychology* 49.4, pp. 764–766 (cit. on p. 47).

Lowe, David G (2004). "SIFT - The Scale Invariant Feature Transform." In: *International Journal of Computer Vision* 60, pp. 91–110 (cit. on p. 70).

Luethje, Fritjof, Dirk Tiede, and Clemens Eisank (2017). "Terrain extraction in built-up areas from satellite stereo-imagery-derived surface models: a stratified object-based approach." In: *ISPRS International Journal of Geo-Information* 6.1, p. 9 (cit. on pp. 18, 90).

Maxant, Jérôme et al. (2013). "Contribution of Pleiades-HR imagery for disaster damage mapping: Initial feedback over Asia, Africa, Europe

or the Caribbean." In: *EARSeL Symposium*. Vol. 33, pp. 491–498 (cit. on p. 1).

Mayer, Nikolaus et al. (2016). "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation." In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4040–4048 (cit. on p. 15).

Meijering, Erik HW, Wiro J Niessen, and Max A Viergever (2001). "Quantitative evaluation of convolution-based methods for medical image interpolation." In: *Medical Image Analysis* 5.2, pp. 111–126 (cit. on p. 37).

Meng, Xuelian, Nate Currit, and Kaiguang Zhao (2010). "Ground filtering algorithms for airborne LiDAR data: A review of critical issues." In: *Remote Sensing* 2.3, pp. 833–860 (cit. on pp. 17, 50, 51).

Meng, Xuelian, Le Wang, et al. (2009). "A multi-directional ground filtering algorithm for airborne LIDAR." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 64.1, pp. 117–124 (cit. on pp. 17, 49–51, 82, 90).

Merkle, Nina et al. (2017). "Exploiting deep matching and SAR data for the geo-localization accuracy improvement of optical satellite images." In: *Remote Sensing* 9.6, p. 586 (cit. on p. 12).

Misra, Prakhar and Wataru Takeuchi (2017). "Digital Surface Model (DSM) Datasets for Built Structure Height Estimation over Indian Cities." In: *International Remote Sensing Symposium*, pp. 1–4 (cit. on p. 90).

Mora, Brice et al. (2014). "Global land cover mapping: Current status and future trends." In: *Land Use and Land Cover Mapping in Europe*. Springer, pp. 11–30 (cit. on p. 1).

Mousa, Yousif Abdul-kadhim, Petra Helmholz, and David Belton (2017). "New DTM extraction approach from airborne images derived DSM." In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 42.1/W1, pp. 75–82 (cit. on p. 90).

Mousa, Yousif Abdul-kadhim, Petra Helmholz, David Belton, and Dimitri Bulatov (2019). "Building detection and regularisation using DSM and imagery information." In: *The Photogrammetric Record* 34.165, pp. 85–107 (cit. on p. 90).

Nex, Francesco et al. (2015). "ISPRS benchmark for multi-platform photogrammetry." In: *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 2, pp. 135–142 (cit. on p. 15).

Orun, Ahmet Bahadir (1994). "A modified bundleadjustment software for SPOT imagery and photography: tradeoff." In: *Photogrammetric Engineering & Remote Sensing* 60.12, pp. 1431–1437 (cit. on p. 13).

Paar, Gerhard and Wolfgang Pölzleitner (1992). "Robust Disparity Estimation in Terrain Modeling for Spacecraft Navigation." In: *IAPR International Conference on Pattern Recognition*. Vol. 11, pp. 738–741 (cit. on p. 42).

Panagiotakis, Emmanouil et al. (2018). "Validation of Pleiades Tri-Stereo DSM in Urban Areas." In: *ISPRS International Journal of Geo-Information* 7.3, p. 118 (cit. on pp. 3, 90).

Perko, Roland (2004). "Computer Vision for Large Format Digital Aerial Cameras." PhD thesis. Austria: Graz, University of Technology (cit. on p. 37).

Perko, Roland, Manuela Hirschmugl, Janik Deutscher, et al. (2019). "Using multiple along and across track Pléiades stereo images for improved digital surface model generation." In: *EARSeL Symposium*. Vol. 39 (cit. on pp. 1, 9, 93).

Perko, Roland, Manuela Hirschmugl, Evelyn Papst, et al. (2016). "Mapping Singapore by Pléiades Stereo Data: Carbon Reporting and more." In: *EARSeL Symposium*. Vol. 36, pp. 170–170 (cit. on p. 1).

Perko, Roland, Wolfgang Koppe, et al. (2018). "Deriving accurate GCPs from TerraSAR-X stereo data over urban areas." In: *EARSeL Symposium*. Vol. 38, pp. 19–19 (cit. on p. 12).

Perko, Roland, Hannes Raggam, Karlheinz Gutjahr, et al. (2011). "Using worldwide available TerraSAR-X data to calibrate the geo-location accuracy of optical sensors." In: *IEEE International Geoscience and Remote Sensing Symposium*, pp. 2551–2554 (cit. on p. 12).

Perko, Roland, Hannes Raggam, Karlheinz Gutjahr, et al. (2014). "Assessment of the mapping potential of Pleiades stereo and triplet data." In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 103–109 (cit. on pp. 1, 2, 9, 83).

Perko, Roland, Hannes Raggam, Karlheinz Gutjahr, et al. (2015). "Advanced DTM generation from very high resolution satellite stereo images." In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. 2, pp. 165–172 (cit. on pp. 1, 10).

Perko, Roland, Hannes Raggam, and Peter M. Roth (2019). "Mapping with Pléiades: End-to-End Workflow." In: *Remote Sensing*. under review (cit. on pp. 9, 93).

Perko, Roland, Hannes Raggam, Mathias Schardt, et al. (2018). "Very High Resolution Mapping with the Pléiades Satellite Constellation." In: *American Journal of Remote Sensing* 6.2, pp. 89–99 (cit. on pp. 1, 3, 8, 12, 13, 15, 16, 93).

Perko, Roland, Mathias Schardt, et al. (2019). "Multiple View Geometry in Remote Sensing: An Empirical Study based on Pléiades Satellite Images." In: *In IEEE International Geoscience and Remote Sensing Symposium* (cit. on pp. 1, 9, 76, 93).

Perko, Roland and Christopher Zach (2016). "Globally Optimal Robust DSM Fusion." In: *European Journal of Remote Sensing* 49, pp. 489–511 (cit. on pp. 15, 16, 47).

Persson, Henrik J. (2016). "Estimation of boreal forest attributes from very high resolution Pléiades data." In: *Remote Sensing* 8.9, p. 736 (cit. on pp. 1, 90).

Persson, Henrik J. and Roland Perko (2016). "Assessment of Boreal Forest Height from WorldView-2 Satellite Stereo Images." In: *Remote Sensing Letters* 7.12, pp. 1150–1159 (cit. on pp. 1, 90).

Persson, Henrik J., Jörgen Wallerman, et al. (2013). "Estimating forest biomass and height using optical stereo satellite data and a DTM from laser scanning data." In: *Canadian Journal of Remote Sensing* 39.3, pp. 251–262 (cit. on p. 1).

Pfeifer, Norbert et al. (2014). "OPALS - A framework for Airborne Laser Scanning data analysis." In: *Computers, Environment and Urban Systems* 45, pp. 125–136 (cit. on pp. 16, 25).

Piermattei, Livia, Mauro Marty, Christian Ginzler, et al. (2019). "Pléiades satellite images for deriving forest metrics in the Alpine region." In: *International Journal of Applied Earth Observation and Geoinformation* 80, pp. 240–256 (cit. on p. 1).

Piermattei, Livia, Mauro Marty, Wilfried Karel, et al. (2018). "Impact of the Acquisition Geometry of Very High-Resolution Pléiades Imagery on the Accuracy of Canopy Height Models over Forested Alpine Regions." In: *Remote Sensing* 10.10, p. 1542 (cit. on pp. 1, 25, 27, 70).

Piltz, Björn, Steven Bayer, and Aanna-Maria Poznanska (2016). "Volume based DTM generation from Very High Resolution Photogrammetric

DSMs." In: *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 41, pp. 83–90 (cit. on p. 18).
Pock, Thomas, Lukas Zebedin, and Horst Bischof (2011). "TGV-fusion." In: *Rainbow of Computer Science*. Springer, pp. 245–258 (cit. on pp. 15, 16, 47).
Poli, Daniela et al. (2013). "Evaluation of Pléiades-1A triplet on Trento testfield." In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 1, pp. 287–292 (cit. on p. 4).
Poli, Daniela et al. (2015). "Radiometric and geometric evaluation of GeoEye-1, WorldView-2 and Pléiades-1A stereo images for 3D information extraction." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 100, pp. 35–47 (cit. on p. 4).
Qayyum, Abdul, Aamir Saeed Malik, and Muhammad Naufal Bin Muhammad Saad (2015). "Evaluation of digital elevation model using rational polynomial coefficient based on HR Pleiades satellite stereo imagery." In: *IEEE International Conference on Signal and Image Processing Applications*, pp. 32–37 (cit. on p. 3).
Qin, Rongjun (2016). "RPC stereo processor (RSP) – A software package for digital surface model and orthophoto generation from satellite stereo imagery." In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. 3, pp. 77–82 (cit. on p. 14).
Qin, Rongjun (2017). "Automated 3d recovery from very high resolution multi-view satellite images." In: *American Society for Photogrammetry and Remote Sensing Annual Conference*, pp. 1–10 (cit. on p. 2).
Raggam, Hannes (2006). "Surface mapping using image triplets." In: *Photogrammetric Engineering & Remote Sensing* 72.5, pp. 551–563 (cit. on pp. 5, 15).
Raggam, Hannes, Alexander Almer, and MF Buchroithner (1993). "Small to Large-Scale Mapping Using Multi-Resolution Image Data." In: *International Archives of Photogrammetry and Remote Sensing* 29, pp. 184–184 (cit. on p. 19).
Raggam, Hannes, Karlheinz Gutjahr, et al. (2010). "Assessment of the stereo-radargrammetric mapping potential of TerraSAR-X multibeam spotlight data." In: *IEEE Transactions on Geoscience and Remote Sensing* 48.2, pp. 971–977 (cit. on p. 12).
Raggam, Hannes, Roland Perko, et al. (2010). "Accuracy assessment of 3D point retrieval from TerraSAR-X data sets." In: *European Conference on Synthetic Aperture Radar*, pp. 1–4 (cit. on p. 12).

Ranftl, Rene, Thomas Pock, and Horst Bischof (2013). "Minimizing TGV-based variational models with non-convex data terms." In: *International Conference on Scale Space and Variational Methods in Computer Vision*, pp. 282–293 (cit. on p. 14).

Reinartz, Peter et al. (2011). "Orthorectification of VHR optical satellite data exploiting the geometric accuracy of TerraSAR-X data." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 66.1, pp. 124–132 (cit. on p. 12).

Rieg, Lorenzo et al. (2018). "Pléiades Tri-Stereo Data for Glacier Investigations - Examples from the European Alps and the Khumbu Himal." In: *Remote Sensing* 10.10, p. 1563 (cit. on p. 1).

Riegler, Gertrud, Simon D Hennig, and Marco Weber (2015). "WorldDEM - A novel global foundation layer." In: *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, pp. 183–187 (cit. on p. 43).

Rothermel, Mathias et al. (2012). "SURE: Photogrammetric surface reconstruction from imagery." In: *LC3D Workshop*, pp. 1–8 (cit. on pp. 14, 43).

Rousseeuw, Peter J and Christophe Croux (1993). "Alternatives to the median absolute deviation." In: *Journal of the American Statistical Association* 88.424, pp. 1273–1283 (cit. on p. 47).

Rumpler, Markus, Andreas Wendel, and Horst Bischof (2013). "Probabilistic range image integration for DSM and true-orthophoto generation." In: *Scandinavian Conference on Image Analysis*. Vol. 7944. Springer LNCS, pp. 533–544 (cit. on pp. 15, 16, 47).

Rupnik, Ewelina, Marc Pierrot-Deseilligny, and Arthur Delorme (2018). "3D reconstruction from multi-view VHR-satellite images in MicMac." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 139, pp. 201–211 (cit. on pp. 2, 14).

Schardt, Mathias et al. (2018). "3D-Pleiades Mapping for IPCC Reporting." In: *Mapping Urban Areas from Space*. Vol. 2 (cit. on p. 1).

Scharstein, Daniel and Richard Szeliski (2002). "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms." In: *International Journal of Computer Vision* 47.1-3, pp. 7–42 (cit. on p. 14).

Scharstein, Daniel, Tatsunori Taniai, and Sudipta N Sinha (2017). "Semi-global stereo matching with surface orientation priors." In: *International Conference on 3D Vision*, pp. 215–224 (cit. on pp. 14, 40).

Schönfelder, Andreas et al. (2017). "Fusion of Point Clouds derived from Aerial Images." In: *OAGM and ARW Joint Workshop*, pp. 128–132 (cit. on pp. 15, 45).
Seki, Akihito and Marc Pollefeys (2017). "SGM-nets: Semi-global matching with neural networks." In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 231–240 (cit. on p. 14).
Shin, Bok-Suk, Diego Caudillo, and Reinhard Klette (2015). "Evaluation of two stereo matchers on long real-world video sequences." In: *Pattern Recognition* 48.4, pp. 1113–1124 (cit. on p. 14).
Sithole, George and George Vosselman (2004). "Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds." In: *ISPRS journal of photogrammetry and remote sensing* 59.1-2, pp. 85–101 (cit. on p. 17).
Sofia, Giulia et al. (2016). "Comparison of Pleiades and LiDAR digital elevation models for terraces detection in farmlands." In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.4, pp. 1567–1576 (cit. on p. 1).
Stein, Michael L (2012). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media (cit. on p. 55).
Steinnocher, Klaus, Roland Perko, and Martin Hofer (2014). "VHR Tri-Stereo Pléiades Satellitenbilder versus Airborne Laserscan Daten - Vergleich auf Basis von abgeleiteten 3D-Gebäudemodellen." In: *Symposium und Fachmesse für Angewandte GeoInformatik*, pp. 33–38 (cit. on p. 1).
Stumpf, André et al. (2014). "Surface reconstruction and landslide displacement measurements with Pléiades satellite images." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 95, pp. 1–12 (cit. on pp. 2, 83).
Stylianidis, Efstratios et al. (2019). "FORSAT: A 3D forest monitoring system for cover mapping and volumetric 3D change detection." In: *International Journal of Digital Earth*, pp. 1–32 (cit. on p. 1).
Tachikawa, Tetushi et al. (2011). *ASTER global digital elevation model version 2-summary of validation results*. Tech. rep. NASA (cit. on p. 43).
Tian, Jiaojiao, Thomas Krauss, and Peter Reinartz (2014). "DTM generation in forest regions from satellite stereo imagery." In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40.1, pp. 401–405 (cit. on p. 18).

Tomasi, Carlo and Roberto Manduchi (1998). "Bilateral filtering for gray and color images." In: *International Conference on Computer Vision*. Vol. 98, pp. 839–846 (cit. on p. 87).

Topan, Hüseyin et al. (2019). "Comprehensive evaluation of Pléiades-1A bundle images for geospatial applications." In: *Arabian Journal of Geosciences* 12.7, p. 223 (cit. on pp. 3, 4).

Toutin, Thierry (2011). "State-of-the-art of geometric correction of remote sensing data: a data fusion perspective." In: *International Journal of Image and Data Fusion* 2.1, pp. 3–35 (cit. on p. 19).

Tuytelaars, Tinne and Krystian Mikolajczyk (2008). "Local invariant feature detectors: a survey." In: *Foundations and trends® in Computer Graphics and Vision* 3.3, pp. 177–280 (cit. on p. 70).

Unger, Markus et al. (2009). "A variational approach to semiautomatic generation of digital terrain models." In: *International Symposium on Visual Computing*, pp. 1119–1130 (cit. on pp. 18, 55).

Vanderhoof, Melanie and Clifton Burt (2018). "Applying High-Resolution Imagery to Evaluate Restoration-Induced Changes in Stream Condition, Missouri River Headwaters Basin, Montana." In: *Remote Sensing* 10.6, p. 913 (cit. on p. 1).

Vivone, Gemine et al. (2015). "A critical comparison among pansharpening algorithms." In: *IEEE Transactions on Geoscience and Remote Sensing* 53.5, pp. 2565–2586 (cit. on p. 4).

Wang, Mi, Fen Hu, and Jonathan Li (2011). "Epipolar resampling of linear pushbroom satellite imagery by a new epipolarity model." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 66.3, pp. 347–355 (cit. on pp. 13, 37).

Warth, Gebhard et al. (2019). "DSM-based identification of changes in highly dynamic urban agglomerations." In: *European Journal of Remote Sensing* 52.1, pp. 322–334 (cit. on p. 1).

Weickert, Joachim and Martin Welk (2006). "Tensor field interpolation with PDEs." In: *Visualization and Processing of Tensor Fields*. Springer, pp. 315–325 (cit. on p. 55).

Wiechert, Alexander and Michael Gruber (2010). "DSM and true ortho generation with the UltraCam-L - A case study." In: *American Society for Photogrammetry and Remote Sensing Annual Conference*, pp. 26–30 (cit. on p. 18).

Yoon, Kuk-Jin and In So Kweon (2006). "Adaptive support-weight approach for correspondence search." In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 4, pp. 650–656 (cit. on p. 14).

You, Yongfa et al. (2018). "Building detection from VHR remote sensing imagery based on the morphological building index." In: *Remote Sensing* 10.8, p. 1287 (cit. on p. 1).

Zabih, Ramin and John Woodfill (1994). "Non-parametric local transforms for computing visual correspondence." In: *European Conference on Computer Vision*, pp. 151–158 (cit. on pp. 14, 39).

Zbontar, Jure and Yann LeCun (2016). "Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches." In: *Journal of Machine Learning Research* 17.1, pp. 2287–2318 (cit. on p. 15).

Zhu, Xiao Xiang et al. (2017). "Deep learning in remote sensing: A comprehensive review and list of resources." In: *IEEE Geoscience and Remote Sensing Magazine* 5.4, pp. 8–36 (cit. on p. 15).