Dissertation

Potential of Very High Resolution Satellite Imagery for 3D Reconstruction and Classification

Ausgeführt zum Zwecke der Erlangung des akademischen Grades einer

Doktorin der technischen Wissenschaften (Dr.techn.)

Unter der Leitung von
Univ.-Prof.Dipl.-Ing. Dr.techn. Norbert Pfeifer

E 120.7
Department für Geodäsie und Geoinformation
Forschungsgruppe Photogrammetrie

Eingereicht an der Technischen Universität Wien
Facultät für Mathematik und Geoinformation

von

Ana-Maria Loghin

Matrikelnummer: 01529410

Wien, am 08. September 2022

....
Unterschrift

Dissertation

Potential of Very High Resolution Satellite Imagery for 3D Reconstruction and Classification

A thesis submitted in fulfilment of the academic degree of
"Doktorin der technischen Wissenschaften (Dr.techn.)"

under the supervision of
Univ.-Prof.Dipl.-Ing. Dr.techn. Norbert Pfeifer

E 120.7
Department of Geodesy and Geoinformation
Research Group Photogrammetry

Research conducted at TU Wien
Faculty of Mathematics and Geoinformation

by
Ana-Maria Loghin

Matriculation number: 01529410

Vienna, September 08, 2022

....
Signature

Supervisor: Prof. Dr. Norbert Pfeifer
Technische Universität Wien
Department of Geodesy and Geoinformation
Research Division Photogrammetry
Wiedner Hauptstraße 8/E120, 1040 Vienna, Austria
E-Mail: Norbert.Pfeifer@geo.tuwien.ac.at

....

Referee: Prof. Dr. Fabio Remondino
Bruno Kessler Foundation (FBK)
Digital Industry Center
3D Optical Metrology (3DOM) Research Unit
Via Sommarive 18, 38123 Trento, Italy
E-Mail: remondino@fbk.eu

....

Referee: Prof. Dr. Martin Rutzinger
University of Innsbruck
Institute of Geography
Remote Sensing & Topographic LiDAR Research
Group
Innrain 52, 6020 Innsbruck, Austria
E-Mail: martin.rutzinger@uibk.ac.at

....

Erklärung zur Verfassung der Arbeit
Author's Statement

Hiermit erkäre ich, dass ich diese Arbeit selbstständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

I hereby declare, that I independently drafted this manuscript, that all sources and references used are correctly cited and that the respective parts of this manuscript including tables, maps and figures - which were included from the manuscript or the internet, either semantically or syntactically -, are made clearly evident in the text and all respective sources are correctly cited.

Ana-Maria Loghin
Hainburgerstraße 24a/24
1030 Wien
Österreich

....
(Unterschrift / Signature)

" Truth is so obscure in these times,
that ...

... unless we love the Truth, we cannot know it ... *"*

Blaise Pascal

Acknowledgements

In this special moments, marked by the end of my PhD research study, I would like to sincerely thank and express my thoughtful gratitude to all people that helped and supported me along this journey.

Firstly, I deeply thank my Supervisor Prof. Norbert Pfeifer for giving me the great opportunity and chance to work and learn about Photogrammetry within the Department of Geodesy and Geoinformation at TU Wien. For me it was a real privilege to be part of the exceptional scientific Photogrammetry Group. His continuous guidance, help, understanding and all the constructive discussions full of valuable ideas and suggestions inspired and helped my research progress. Without such support, this work would not have been possible. *Dear Norbert, I feel privileged that I have had the chance to be under Your supervision! Thank You!*

I specially thank Dr. Johannes Otepka-Schremmer for his help, collaboration and guidance throughout all this period, but also for the great discussions and reviews that were always motivating me to a deeper understanding of the research result interpretation and further investigations.

Particular thanks to Wilfried Karel for his time and patience in helping and explaining me the programming side, to Dr. Camillo Ressl who influenced my view on photogrammetry, to Dr. Gottfried Mandlburger for his inspirational research and great discussions, to Dr. Hollaus Markus for giving me the first insights into the OPALS software, to the Secretary, OPALS, and IT teams for their permanent kind assist and technical support.

The support from many of my colleagues was very important for me during this work. Therefore I would like to thank Sebastian Mikolka-Flöry, Markus Pöchtrager, Nan Li, Anna Iglseder, Michael Wimmer, Taskin Özkan, Moritz Bruggiser, Livia Piermattei for the many interesting discussions and the nice time spent together.

Sincerely thanks to Dr. Oniga Ersilia from the Technical University of Iasi, who opened up for me the door of this great opportunity, to come and study in Vienna. Deeply and sincerely thanks to the beautiful roumanian orthodox community of people and friends in Vienna for the valuable moments spend together. I also thank my friend, Alexandra Bojor, who encouraged me to apply for this PhD position.

Last but not least, I thank with all my heart my family: my mother, sisters and all close ones, for believing in me and for their continuous help, confidence, patience, and love with which they surrounded me and helped me a lot to pursue my research. I deeply thank my father, who even if he is not physically among us, I always felt him close, I felt his continuous help and encouragement.

Thank God for my entire Life, Everyone and Everything within it!

Abstract

Very High Resolution (VHR) satellite sensors with their efficient worldwide acquisition capabilities are a valuable source for geospatial applications requiring large-scale coverage and high temporal resolutions, such as: disaster monitoring, urban management, change detection, navigation support, environmental and hydrological modelling. However, the purely two-dimensional image information is often not sufficient for assessing different specific situations and therefore a three-dimensional reconstruction of the depicted scene is necessary. By combining the 2D image content with the 3D reconstructed geometry, improved baselines for various applications can be achieved, leading to a better understanding of the observed scene. This is possible with the today's high agile sensors able to acquire multi-view satellite image collections over the same area of interest during a single pass. While many research papers address the spectral 2D content of optical satellite images, the 3D capability exploration topic is significantly reduced. Given the low amount of research in this direction, the aim of this dissertation is to increase and expand the knowledge regarding the 3D potential of VHR satellite imagery. The focus falls on the automatic 3D information extraction and classification of natural and developed areas from stereo/tri-stereo scenes of VHR satellite images from Pléiades and WorldView-3 in order to highlight and to get a deep insight into the potential and limitations of VHR sensors.

The research in this dissertation focuses on both methodology developments and descriptive aspects for improving and better understanding the actual state of the achievable 3D information from VHR satellite imagery, specifically associated with following scientific objectives: (a) 3D satellite-based point cloud reconstruction and accuracy analysis of elevation models; (b) satellite image geometry correction based on high resolution elevation models; (c) orthophoto quality improvement; (d) classification performance analysis; and (e) potential for 3D reconstruction of small, isolated objects. A side focus is also paid on the description of an end-to-end photogrammetric workflow for 3D information extraction from VHR satellite imagery and on object visibility analysis based on satellite image acquisition geometry parameters.

Practical examples for 3D information extraction are analysed in four distinct areas with various topographic characteristics and land coverage: rural, urban, open agricultural fields, grasslands, forests, and mountainous areas. Conducted experiments reveal possible vibrations of the satellite sensor (jitter effect), caused by the fast rotation from one viewing direction to another, which could not be captured by the provided Rational Polynomial Coefficients (RPCs). A new method is introduced that optimizes the geometry of satellite images based on the back-projection of actual and reference elevations, by means of bias-corrected RPCs. Besides methods for improving the geometric accuracy and quality of satellite-based elevation models and orthophotos, special attention was accorded to the finest details that can be mapped in the photogrammetrically derived elevation models.

By addressing the 3D achievable information extraction from Pléiades and WordView-3 stereo/tri-stereo imagery, these contributions bring new insights to the research topic and can help to increase and expand the knowledge regarding the 3D potential of VHR satellite imagery.

Kurzfassung

Hochauflösende Satellitensensoren mit ihrem effizienten weltweiten Aufnahmemöglichkeiten sind eine wertvolle Quelle für Geodatenanwendungen, die eine großflächige Abdeckung und hohe zeitliche Auflösungen erfordern, wie zum Beispiel: die Überwachung von Umweltkatastrophen, Stadtplanung, Veränderungsdetektion, Navigationsunterstützung, Umwelt- und hydrologische Modellierung. Allerdings reicht die alleinige zweidimensionalen Bildinformationen zur Beurteilung unterschiedlicher spezifischer Situationen oft nicht aus. Daher ist eine dreidimensionale Rekonstruktion der abgebildeten Szene erforderlich. Durch die Kombination des 2D-Bildinhalts mit der rekonstruierten 3D-Geometrie kann eine verbesserte Basis für verschiedene Anwendungen erreicht werden. Das führt zu einem besseren Verständnis der beobachteten Szene. Dies ist mit den heutigen hochflexiblen Sensoren möglich, die Satellitenbilder aus mehreren Ansichten über denselben Interessenbereich während eines einzigen Durchgangs aufnehmen können.

Während sich viele Forschungsarbeiten mit dem spektralen 2D-Inhalt optischer Satellitenbilder befassen, ist das Thema der 3D-Untersuchungen erheblich kleiner. Angesichts des geringen Forschungstätigkeiten in dieser Richtung zielt diese Arbeit darauf ab, das Wissen über das 3D-Potenzial von hochauflösenden Satellitenbildern zu erhöhen und zu erweitern. Das Ziel dieser Dissertation ist die automatische 3D-Informationsextraktion und Klassifizierung von natürlichen und bebauten Gebieten aus Stereo/Tri-Stereo-Szenen von Pléiades und WorldView-3 Satellitenbildern, um das Potenzial und die Grenzen hervorzuheben und einen tiefen Einblick von hochauflösenden-Sensoren zu erhalten.

Die Arbeit der Dissertation fokussiert sowohl auf methodische Entwicklungen als auch auf beschreibende Aspekte zur Verbesserung und zu einem besseren Verständnis des aktuellen Zustands der erreichbaren 3D-Informationen aus hochauflösenden Satellitenbildern. Diese sind insbesondere mit den folgenden wissenschaftlichen Zielen verbunden: (a) 3D-satellitenbasierte Punktwolkenrekonstruktion und Genauigkeitsanalyse von Höhenmodellen; (b) Satellitenbild-Geometriekorrektur basierend auf hochauflösenden Höhenmodellen; (c) Verbesserung der Orthofotoqualität; (d) Analyse der Klassifikationsleistung; und (e) Potenzial für die 3D-Rekonstruktion kleiner, isolierter Objekte. Ein Nebenfokus liegt auf der Beschreibung eines photogrammetrischen Arbeitsablaufs für die 3D-Informationsextraktion aus hochauflösenden Satellitenbildern und auf der Objektsichtbarkeitsanalyse basierend auf den Geometrieerfassungsparametern.

Praktische Beispiele für die Extraktion von 3D-Informationen werden in vier unterschiedlichen Gebieten mit unterschiedlichen topografischen Eigenschaften und Landbedeckung analysiert: ländliche, städtische, offene landwirtschaftliche Felder, Grasland, Wälder, und Berggebiete. Die durchgeführten Experimente zeigen vermutliche Vibrationen des Satellitensensors (Jitter-Effekt), die durch die schnelle Drehung von einer Blickrichtung in die andere verursacht werden und von den mitgelieferten rationalen Polynomkoeffizienten (RPC) nicht erfasst werden konnten. Es wird eine neue Methode eingeführt, die die Geometrie von Satellitenbildern basierend auf der Rückprojektion von reale und Referenzhöhen unter Verwendung von Bias-korrigierten RPCs optimiert. Neben Methoden zur Verbesserung der geometrischen Genauigkeit und Qualität von satellitengestützten Höhenmodellen und

Orthofotos wurde besonderes Augenmerk auf feinste Details gelegt, die in den photogrammetrisch abgeleiteten Höhenmodellen abgebildet werden können.

Indem sie die erreichbare 3D-Informationsextraktion aus Pléiades- und WordView-3 Stereo/Tri-Stereo-Bildern ansprechen, bringen diese Beiträge neue Einblicke in das Forschungsthema und tragen dazu bei, das 3D-Wissenspotenzial von hochauflösenden Satellitenbildern zu erhöhen und zu erweitern.

Rezumat

Sateliții de mare rezolutie cu capabilitățile lor eficiente de achiziție la nivel global reprezintă o sursă valoroasă pentru aplicațiile geospațiale care necesită o acoperire pe scară largă și rezoluții temporale ridicate, cum ar fi: monitorizarea dezastrelor naturale, managementul urban, detectarea schimbărilor în timp, suport pentru navigație, modelarea de mediu și hidrologică. Cu toate acestea, informațiile pur bidimensionale ale imagininilor satelitare nu sunt adesea suficiente pentru a evalua diferite situații specifice și, prin urmare, este necesară o reconstrucție tridimensională a scenei reprezentate. Prin combinarea conținutului imaginii 2D cu geometria reconstruită 3D, pot fi obținute linii de bază îmbunătățite pentru diverse aplicații, ceea ce duce la o mai bună înțelegere a scenei observate. Acest lucru este posibil prin senzorii foarte flexibili de astăzi, capabili de a obține colecții de imagini satelitare cu mai multe achizitii pentru aceeași zonă de interes în timpul unui singur zbor. În timp ce multe lucrări de cercetare abordează conținutul spectral 2D al imaginilor optice satelitare, subiectul de explorare a capacității 3D este redus în mod semnificativ. Având în vedere volumul redus de cercetări în această direcție, această lucrare își propune să crească și să extindă cunoștințele privind potențialul 3D al imaginilor satelitare de mare rezoluție. Scopul acestei disertații este de a extrage în mod automat informațiile 3D și de a clasifica arealele naturale și artificiale prin folosirea de imagini stereo/tri-stereo provening de la sateliții Pléiades și WorldView-3 pentru a evidenția și a obține o perspectivă profundă asupra potențialului și limitărilor acestor senzori.

Cercetarea abordată în această disertație se concentrează atât pe dezvoltarea metodologiei, cât și pe aspecte descriptive pentru îmbunătățirea și înțelegerea mai profundă a situației actuale în ceea ce privește gradul de informații 3D ce pot fi extrase din imaginile satelitare de mare rezoluție, problematici asociate în mod specific cu următoarele obiective științifice: (a) reconstrucția norilor de puncte 3D și analita acurateței modelelor de elevație; (b) corecția geometriei imaginilor satelitare pe baza modelelor de mare rezoluție; (c) îmbunătățirea calității modelului ortofoto digital; (d) analiza performanței clasificării norilor de puncte; și (e) potențialul pentru reconstrucția 3D a obiectelor mici, izolate. O atenție secundară este, de asemenea, acordată pe descrierea unui flux de lucru fotogrammetric end-to-end pentru extragerea de informații 3D din imaginile satelitare de mare rezoluție și pe analiza vizibilității obiectelor bazată pe parametrii geometrici de achiziție ai imaginilor satelitare.

Exemplele practice analizate pentru extragerea informațiilor 3D sunt localizate în patru zone distincte cu diferite caracteristici topografice și utilizare a terenului: rural, urban, câmpuri agricole deschise, pajiști, păduri și zone muntoase. Experimentele efectuate relevă vibrații ale sateliților (efect de fluctuație), cauzate de rotația rapidă dintr-o direcție de vizualizare în alta, care nu au putut fi captate de parametrii RPC de orientare furnizați. Așadar, este introdusă o nouă metodă care optimizează geometria imaginilor satelitare pe baza retro-proiecției elevațiilor reale și de referință, prin intermediul parametrilor RPC corectați. Pe lângă metodele de îmbunătățire a acurateței geometrice și a calității modelelor de elevație și a modelelor ortofoto, s-a acordat o atenție deosebită celor mai fine detalii care pot fi reconstruite în modelele de elevație satelitare.

Prin abordarea tematicii în ceea ce privește gradul realizabil de informații 3D ce pot fi extrase pe baza imaginilor stereo/tri-stereo provenind de la sateliții Pléiades și WordView-3, aceste contribuții aduc noi perspective asupra temei de cercetare și ajută la creșterea și extinderea cunoștințelor cu privire la potențialul 3D al imaginilor satelitare de mare rezoluție.

Abbreviations

ALS	Airborne Laser Scanning
CART	Classification and Regression Trees
CBM	Cost Based Matching
CPs	Check Points
DEM	Digital Elevation Model
DIM	Dense Image Matching
DSM	Digital Surface Model
DTM	Digital Terrain Model
FBM	Feature Based Matching
GCPs	Ground Control Points
GIS	Geographic Information Systems
GSD	Ground Sampling Distance
LiDAR	Light Detection And Ranging
LSM	Least Squares Matching
nDSM	normalized Digital Surface Model
OSM	Open Street Map
RFM	Rational Function Model
RMSE	Root Mean Square Error
RPC	Rational Polynomial Coefficient
RTK	Real Time Kinematic
TPs	Tie Points
VHR	Very High Resolution

Abbreviations

Contents

Contents

xii

1

Introduction

1.1 Motivation

Surveying the planet we live on has been always an ongoing effort that met multiple evolution stages through time. From ropes and pegs in Antiquity, theodolite in the 15th century, total stations and GPS base stations at the end of 20th century, to today's modern technologies of 3D laser scanning, mobile mapping, drones and satellite systems, the mission of any surveying activity was always to make sense of the world around us.

Image matching of aerial or satellite images and airborne Light Detection And Ranging (LiDAR) technology serve as main techniques for the acquisition of three dimensional geospatial information, used for mapping and 3D modelling of large surface areas. With their advantages and limitations, both technologies have been exploited for their capabilities on the production of large area high resolution Digital Surface Models (DSMs). To guarantee reasonable analyses in the multiple applications, surface models with high quality are needed. Airborne Laser Scanning (ALS) active systems are the primary data source for accurate Digital Elevation Model (DEM) extraction of high spatial resolution (Contreras et al. 2017, Habib et al. 2005) with reported height accuracy below 25 cm (F. J. Aguilar et al. 2008, X. Liu et al. 2007). The great advantage of this technology is the ability of capturing directly three-dimensional information in object space, based on the flight of a light signal between the sensor and the target (Vosselman et al. 2010). Nevertheless, the major limitations of ALS data are related to access restrictions and low temporal resolution. On the other hand, multi view aerial imagery is the traditional and cost efficient method to acquire high resolution elevation data over large surface areas. Depending on the Ground Sampling Distance (GSD) of the used sensor, the best vertical accuracies of airborne photogrammetrically derived DEMs vary between 0.44 and 2 GSD (Bühler et al. 2012, Hobi et al. 2012, Hoffmann et al. 2000, W. Hu et al. 2008, Ressl, Brockmann, et al. 2016).

Although LiDAR and aerial imagery provide dense and accurate 3D surface information, both techniques are limited by the small area coverage, low acquisition frequency due to restrictions in data acquisition, security reasons and authorisation. In contrast, satellite images may capture any point of the world at regular time intervals. Although satellite sensors are inferior to airborne cameras in terms of ground resolution as well as configuration options for multi-view images, the major advantages of satellite imagery are the large area coverage within a very short time (seconds), the worldwide availability, the possibility of remote recording of areas of interest that are difficult to reach or impossible to access due to airspace

legal restrictions (crises, natural disasters, remote regions, war zones). In addition, characterized by a high temporal resolution (few days interval), satellites allow a cost-effective recording of several epochs with repetition cycles of a few days (used for monitoring, change and anomaly detection). Furthermore, the rich semantic information contents of the satellite images enhances land cover classification applications, object identification, change detection and further visual interpretation and analyses. As a result, satellite-based sensors can cover strategic, planning and safety-relevant aspects which can not be appropriately processed with airborne platforms, nor can be covered by other measuring techniques.

Earth observation by means of satellite data offers on the one hand, the possibility of observing interest areas at regular intervals (i.e. high temporal resolution), and on the other hand, the ability to analyze areas outside countries borders in real time. The availability of satellite data together with the efficient data acquisition over large areas has led to an upward trend in many applications which require large-scale coverage and high temporal resolutions, such as disaster monitoring, urban management, change detection, hydrological modelling. These applications usually require up-to-date elevation models and corresponding maps at a large scale, a fact that increases the need for developing automatic workflows based on remote sensing data.

Due to the further development of remote sensing systems, spatial resolutions in the range of less than one meter can now also be achieved through satellite-based recordings. However, the purely two-dimensional image information is often not sufficient for assessing different specific situations and therefore a three-dimensional reconstruction of the depicted scene is necessary. This is possible by viewing the same area from different observation points (e.g. from different positions of the satellite in the orbit). Current optical earth observation satellites are able to collect multi-view images of the same area during a single pass, and thus optical satellite imagery became an important source for 3D information extraction. By combining the two-dimensional image content with the 3D reconstructed geometry, improved baselines for various applications can be achieved, including automated classification, that leads to a better understanding of the observed scene. In its turn, classification leads to further complex investigations like change detection and 3D city modelling.

For more than thirty years, the interest of the scientific community has turned to the great potential of the civilian satellite sensors for providing 3D geospatial information over large areas in a timely and cost-effective manner (d'Angelo and Kuschk 2012). Due to the new dense matching techniques and their improved spatial resolution, satellite imagery is becoming an important data source for 3D information extraction and DSM derivation (Xie et al. 2019). With a large area coverage, photogrammetrically-derived DEMs have a wide range of applications in topography, cartography, geomorphology, engineering, land planning, disaster management and 3D surface modelling.

Beginning with the year 1986 when SPOT - the first satellite providing stereo-images, with a panchromatic GSD of 10 m - was launched (Toutin and Cheng 2002), the optical satellite industry has been experiencing a continuous development. Among the most important sensor characteristics that were analyzed and improved in time are those referred to the capability to describe a scene in terms of geometrical detail (GSD), radiometric sensitivity, spectral content and temporal frequency of

the acquisition.

Today more and more space sensors are available that acquire not only stereo but also tri-stereo satellite imagery. The possibility of DEM extraction from satellite stereoscopic images started in 1986. Thanks to the continuous development of the remote sensing systems, spatial resolutions in the range of less than one meter can be nowadays achieved by means of satellite-supported images. Very high resolution optical sensors are able to acquire images comparable with those provided by aerial imagery, in terms of high GSD values (0.30 m). The generation of high and Very High Resolution (VHR) commercial space imaging systems for DEM generation started in September 1999, with the launch of IKONOS with 1 m and 4 m GSD panchromatic and multispectral, respectively (Sadeghian et al. 2001). Among the VHR optical satellites providing along- and across-track stereo, the following systems can be mentioned: KOMPSAT-2 (1 m), Gaofen-2 (0.8 m), TripleSat (0.8 m), EROS B (0.7 m), KOMPSAT-3 (0.7 m), Pléiades 1A/1B (0.5 m), SuperView 1-4 (0.5 m), GeoEye-1 (0.5 m), WorldView-1/2 (0.5 m), and WorldView-3/4 (0.3 m, WorldView-4 retired in January 2019). The journey of high resolution satellites is expected to continue with a greater number of advanced sensors with higher spatial and temporal resolutions. Such expanded high resolution constellations are announced and already started the launch in 2021, like the Pléiades NEO 3 and 4 from Airbus in April and August, the first two satellites in the Pléiades NEO constellation of four agile satellites, offering 30 cm resolution (Airbus 2020). Phased at 90° and operated on the same orbit, the satellites are able to visit any point on the globe at least twice daily. Their high agility and reactive tasking ability allows a faster coverage and acquisitions up to 30 - 40 minutes following request, being able to respond to the most critical situations in almost near-real time (ESA 2021). The next competitive high-performance evolution will come from the american Maxar Technologies Inc. company with the WorldView-Legion constellation of six satellites able to collect sub 30 cm resolution imagery (Maxar 2020). The launch of the first two satellites is planned for middle March 2022. These satellites will replace imaging capability currently provided by DigitalGlobe's WorldView-1, WorldView-2 and GeoEye-1 Earth observation satellites. The ability to highly revisit the most in-demand areas (like for e.g. the changing areas on Earth) up to 15 times per day will allow a better critical inform and decision-making for time-sensitive applications. The advanced constellation will offer high-agility point collections for monitoring regions of interest and large-area collections for mapping missions, enabling a constant monitoring, near-real time change detection and timely analysis at scale (eoPortal 2021). A future launch is planned for WorldView-Scout satellites from Digital Globe (a network of six or more small satellites with 80 cm resolution) that will be able to revisit key areas of the globe 40 times a day (Gunter 2020).

Some of the key players operating the commercial satellite imaging market include Maxar Technologies Inc. (DigitalGlobe and MDA Holdings Company), Environmental Systems Research Institute (ESRI), BlackSky Global LLC, Harris Corporation, Planet Labs Inc., and Space Know in USA, European Space Imaging in Germany, Airbus Defence and Space in France, ImageSat International N.V. in Israel, Galileo Group in Australia. Key players are adopting numerous strategies like product development, collaborations, partnership, acquisition, agreement, product launch, and business expansion to stay competitive in the commercial satellite imaging market (Allied-Market-Research 2020).

Chapter 1. Introduction

In this dissertation, the focus is on exploring the capabilities and limitations of two VHR satellite sensors, Pléiades and WorldVew-3, with respect to 3D information extraction and classification. The new generation of Earth observation satellites are characterized by an increased acquisition capacity and the possibility of collecting multiple images of the same area from different viewing angles during a single pass (Astrium 2012, DigitalGlobe 2012). The Pléiades dual system is operated by the french space agency CNES - Centre National d'Etudie Spatiales), but has been developed in a consortium with the participation of the following european countries: Italy, Belgium, Spain, Sweden, and Austria. Active since December 2011, the constellation is capable of acquiring imagery of any point on the globe in under 24 hours and is specifically designed to deliver high resolution optical images for defense, civil security, emergency response, and mapping applications (*https://pleiades.cnes.fr/en/PLEIADES/GP_systeme.htm*).

The WorldView-3 Earth observation satellite, launched in August 2014, is the sixth satellite in orbit owned by the American company DigitalGlobe, joining Ikonos (which was launched in 1999), QuickBird (in 2001), WorldView-1 (in 2007), GeoEye-1 (in 2008), and WorldView-2 (in 2009). Near the newly launched Pléiades Neo satellites in 2021, WorldView-3 offers nowadays one of the highest spatial resolution among the commercially available sensors with stereo capabilities, having a wide range of applications in agriculture mapping, environmental monitoring, energy industry (oil and gas), crisis management, and mining (*http://worldview3.digitalglobe.com/*).

Both Pléiades and WordView-3 satellite imaging systems have similar configuration characteristics. Equipped with a combination of linear Charge Coupled Device (CCD) array sensors, the scanners record one line of the image simultaneously and perpendicular to the flight direction. The agile observation system based on the pushbroom principle is also called along track scanner. This allows stereo recordings along the flight direction ("along track stereo") with a single observation line (as opposed to "across track stereo", Nikolakopoulos et al. 2005). Moreover, the Pléiades and WorldView-3 VHR optical satellite sensors are able to collect not only stereo but tri-stereo images of the same area during a single flight path, from different viewing angles along the track: forward, close to nadir and backward. This ability has many advantages with regard to point cloud completeness and DSM derivation (Panagiotakis et al. 2018, Piermattei, Marty, Karel, et al. 2018, Rupnik et al. 2018) especially met in mountainous or urban areas, where the possibility of hidden items or regions (in between two high mountains or around buildings) is higher. Hence, the tri-stereo acquisition mode mitigates the risk of occlusions by providing better area coverage (Figure 1.1).

The reconstructed 3D point cloud, or in less complex 3D situations a textured 2.5D elevation model, represent the final result for some applications (e.g. visualizations). For complex tasks, however, it is often necessary to use 3D information (point clouds) with semantic information, where the semantic distribution of the point clouds is usually referred to as classification. In recent years there has been much progress in machine learning and many different classification algorithms are available (Grilli et al. 2017). In this work, to analyse and understand the quality of the satellite-driven 3D scenes for semantic segmentation, the Pléiades and WorldView-3 point clouds were classified by using a supervised machine learning algorithm employing decision trees.
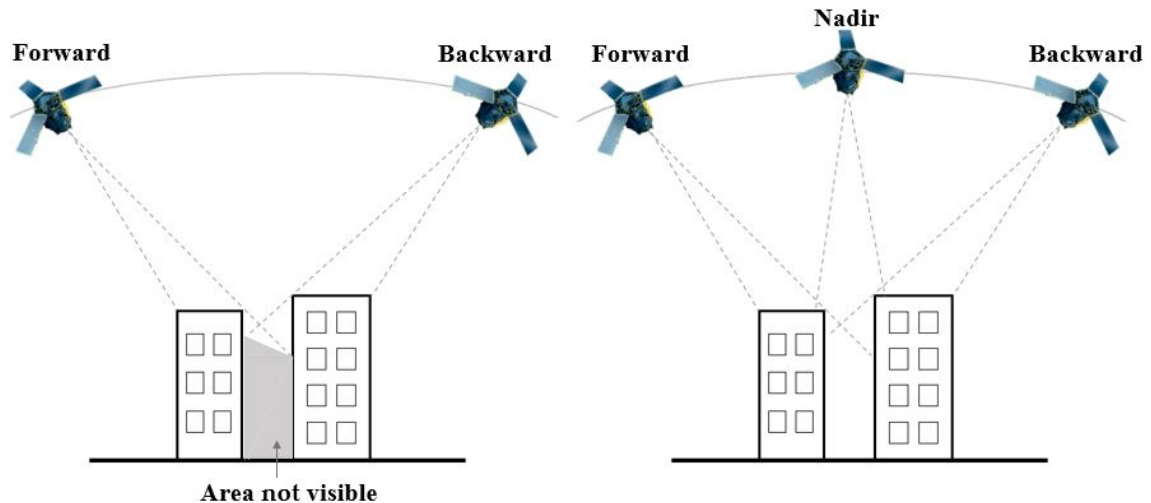
Figure 1.1: Stereo and tri-stereo cover capabilities over urban areas.

In the last four decades, according to the Web of Science global citation database, there has been an increase in the number of publications focusing on optical satellite imagery applications. A search query of articles specifically containing the words "satellite" in the title and "imagery/image/images" and "optic" in the abstract during the last 36 years (1986 - 2022) returned a total of 1321 articles. The year 1986 was chosen as a reference time, since this was the starting point for DEM extraction from stereo satellite images (2002).

It is evident that since 1986 there had been a continual increase in the number of published research papers in the optical satellite imagery research topic (Figure 1.2). This indicates that the current topic, as a research domain, has been gaining more importance to specialists over the time. A higher increase is visible starting with 2015 year, which actually shows that satellite images are getting more and more used for multiple applications. Nevertheless, the number of publications that address the 3D capabilities and elevation model extraction from stereo/tri-stereo acquisitions is clearly smaller (only 20% from the total number) in comparison with articles exploring the spectral 2D content of the images. For example, in 2021 only 30 articles address the 3D potential while 160 research papers are dedicated to 2D image analysis.

Given the low amount of research regarding the 3D capabilities of multi-view satellite images, this dissertation aims to increase and expand the knowledge regarding the 3D capabilities of very high resolution satellite imagery. The need of the performed investigations started from real problem applications that require optimal, fast and reliable solutions. First of all, the creation of up-to-date maps containing accurate information play an essential role especially for safety-critical applications (humanitarian and military operations) like environmental and natural catastrophes, technological and industrial accidents. In case of a crisis situation, for the planning and implementation of an evacuation operation from a specific area (e.g. airport), a 3D model is needed within 24 hours or even less. Additionally, detailed classified 3D models allow to analyze coverage possibilities, to determine the dead angles of enemy positions, to visualize the impact areas of their own operational forces, to display landing and take-off places as well as approximation possibilities. Hence,
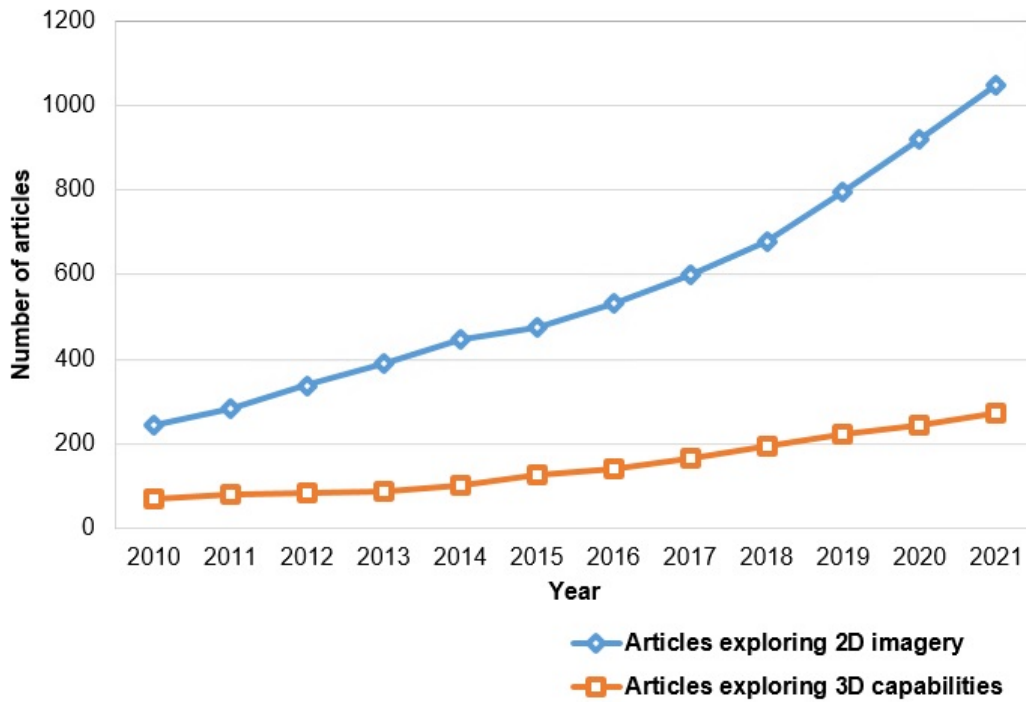
Figure 1.2: Dynamics of optical satellite imagery- related papers.

an automatically workflow for reconstructing 3D scenes from satellite images represent a promising data source for up-to-date map creations. Moreover, for surveying projects abroad, there are always problems for classical flights due to political and military reasons (permits). In such cases, the use of satellite images could be a cost-effective and rapid alternative for obtaining 3D information with the required quality (private communication with military personnel).

The aim of this dissertation is the automatic 3D reconstruction and classification from stereo/tri-stereo very high-resolution satellite images of Pléiades and WorldView-3 in order to get a deep insight into the potential and limitations of satellite imagery for automatic 3D reconstruction in the context of various applications. For instance, although building reconstruction from stereo satellite images in 3D urban modelling is being addressed in some studies (Xu et al. 2020, Kim et al. 2007, Duan et al. 2016, Yi et al. 2019), to the best of our knowledge, no research for single, isolated objects was conducted so far. Hence, besides the geometric accuracy analysis of satellite image-driven products, special attention was accorded to the finest details that can be automatically reconstructed and identified in the photogrammetrically derived elevation models. The objective of this analysis was to inspect how far can we go at the smallest level and where are the limits of VHR satellite imagery with regard to the 3D reconstruction of single objects. For this, isolated buildings, trees and vehicles were selected. Only static parked cars were considered, since moving objects can not be reconstructed through photogrammetric processing. Moreover, through this investigation were defined the conditions that single isolated objects need to fulfill to be automatically reconstructed.

Even if the number of high and very high resolution (VHR) optical satellite sensors is continuously growing, not all high-resolution satellite imaging cameras have a sufficient and stable calibration in time. Due to their high agility in rotation, a quick

change in viewing direction can lead to satellite attitude oscillation, causing image distortions and thus affecting image geometry and geo-positioning accuracy. This dissertation also focuses on a thorough analysis for evaluating and improving the accuracy of elevation models that are obtained from Pléiades tri-stereo imagery and to find the possible sources for existing systematic errors in the DSMs. A principal objective is to develop a new method based on vendor provided Rational Polynomial Coefficients (RPCs) and a high resolution elevation model for improving the geometry of (Pléiades) VHR satellite images. The proposed approach is based on the re-projection of regularly distributed 3D ground points from object in image space, to detect and estimate the periodic distortions of Pléiades tri-stereo imagery caused by satellite attitude oscillations. After applying the periodic distortion compensation in all three images, the existing systematic elevation discrepancies from the derived elevation models is successfully removed and their overall accuracy improved.

With this dissertation we want to explore the potential and capabilities versus the limitations (borders) of VHR satellite imagery in different circumstances and input conditions. Therefore distinct areas of interest with various topographic characteristics and land use (land coverage) were selected: urban areas, characterized by a high activity and city dynamics; rural and land landscapes with open fields; and mountainous areas with high peaks, valleys and glaciers. For these specific areas, different experiments were conducted: satellite image orientation based on the Rational Function Model (RFM) refined by Ground Control Points (GCPs) and Tie Points (TPs), accuracy analysis of photogrammetrically derived DSMs, geolocation improvement, image correction model computation, orthophoto derivation, supervised classification of satellite derived 3D point clouds, automatic height estimation of single, isolated objects, and visibility analysis based on satellite viewing geometry.

By addressing the 3D achievable information content extraction from stereo/tri-stereo imagery of the very high resolution Pléiades and WordView-3 sensors, the current thesis brings knowledge and new insights to the research topic.

1.2 Related work

Literature on VHR Satellite Image Orientation and Attitude Oscillation

Generally, an accurate derivation of spatial and descriptive information from imagery requires a careful calibration of the photogrammetric system, referring to a precise computation of both interior and exterior orientations. The exterior orientation is defined by two important parameters, that can affect the geometric performance of high resolution satellites: camera position and attitude. While camera position is determined as a function of time by the on-board global position system receiver and satellite ephemeris, the platform attitude is merely obtained from the star-trackers and gyros onboard, also as a function of time (Tong, Ye, et al. 2014; Wang et al. 2016).

In order to guarantee a high geometric quality of geospatial information products (such as digital elevation models or orthophoto maps), a precise sensor orientation is required. One of the most critical and important issues in the photogrammetric processing chain of high resolution satellite imagery is the improvement of geopositioning accuracy, a subject which has been continuously studied for years (Fraser,

Dial, et al. 2006; Y. Zhang, Zheng, et al. 2014; Y.-h. Jiang, G. Zhang, Tang, D.-r. Li, et al. 2015; Oh et al. 2015; Shen, Q. Li, et al. 2017; Dong et al. 2020). In satellite imagery, sensor orientation models are establishing the functional geometric relationships between object and image space, or vice versa. The two most widely used imaging geometry models in photogrammetry and remote sensing are the physical and generalized sensor models (Y. Hu et al. 2004; Daniela Poli and Toutin 2012; Toutin 2011). Although physical sensor models explicitly reflect the physical reality by rigorous geometric relationships between 2D image space and 3D object space, they are still not widely used in practice, since their sensor-dependent parameters are kept confidential by some commercial satellite image vendors (Tong, S. Liu, et al. 2010). Because of its simplicity and generality, for more than two decades, the RFM is the most popular generic sensor model used in the field. Defined as a ratio of two cubic polynomials with 80 coefficients, the rational functions have proven to be a feasible alternative solution for replacing the rigorous physical sensor model, due to their high precision in geo-positioning accuracy (Gene Dial et al. 2002; M. A. Aguilar, F. J. Aguilar, Mar Saldaña, et al. 2012; Alkan et al. 2013; Fraser and Hanley 2005; Daniela Poli and Toutin 2012; Tao, Y. Hu, and W. Jiang 2004). However, because of inaccurate measurements of satellite orbits and attitudes, RPCs provided by the commercial satellite image vendors may not always realize with exactness the real imaging process. There exist systematic biases between the coordinates derived by using the vendor provided RPCs and their true ones, resulting errors in image space which range typically from several to tens of pixels (Grodecki et al. 2002; Fraser and Hanley 2003; Noguchi et al. 2004; Teo 2011). Therefore, there is a clear need to correct the existing systematic biases in the RPCs, in order to meet the requirements of high-precision topographic mapping (Shen, B. Liu, et al. 2017). In this research direction, many studies have been conducted on a variety of high resolution satellite images, such as: Quick Bird (Hong et al. 2015; Tong, S. Liu, et al. 2010), Ikonos (Grodecki et al. 2003; Y. Zhang, Lu, et al. 2011), GeoEye (M. A. Aguilar, F. J. Aguilar, Mar Saldaña, et al. 2012; M. A. Aguilar, Mar Saldana, et al. 2013), WorldView 1/2 (Alkan et al. 2013; Teo 2011). Some commonly and widely used RPC-bias correction models are: the translation model, the shift and drift model, the affine transformation model, the conformal transformation model, and the quadratic polynomial model. Improved bias-correction models were developed specific to the chinese Ziyuan-3 sensor, like the bias correction with cubic splines (J. Cao et al. 2017), thin-plate smoothing spines (Shen, B. Liu, et al. 2017), or by local polynomial modelling (Shen, Q. Li, et al. 2017).

Modern satellite imaging systems are pushbroom sensors, equipped with a combination of linear Charge Coupled Device (CCD) array sensors. In flying direction, images are generated by the movement and rotation of the satellite; this means that image geometry directly depends on the relative accuracy of sensor orientation angle and therefore, it can be considered as a function of the exterior orientation (Jacobsen et al. 2008). Regularly, satellite cameras are calibrated before launch; but it cannot be guaranteed that the geometry is not affected by the extreme acceleration during launch and the conditions in space (temperature influences and desiccation effects), requiring an in-flight calibration (Jacobsen et al. 2008; Kornus et al. 1999). Companies and governmental institutions operating optical satellite sensors, usually calibrate their systems in laboratory or in space (Dial et al. 2004), but this is limited by the attitude data quality, which is affected by the permanent satellite rotations

for acquiring the planned ground area (Jacobsen 2018; Tong, Ye, et al. 2014; J. Pan et al. 2017). Usually, for the classical satellites the viewing direction in relation to the orbit was nearly constant during imaging, but this is not the case for the newly very flexible satellites like for e.g. Pléiades. Images can be collected also by scanning against or across the movement in the orbit, and satellites might show vibrations (jitter effect) caused by the fast rotation from one viewing direction to another. Especially for tri-stereoscopic coverage from the same orbit, fast pitch rotations are needed for the switch from forward view to the nadir- and backward-view. Such rotations require a powerful angular acceleration, followed by a negative angular acceleration, which may cause an oscillation during imaging. These vibrations need to be measured by the gyros component, but actually not all gyros are able to record the oscillation frequency. Hence, a limited accuracy of the view direction is causing systematic image errors in relation to the used mathematical model of geometric reconstruction. Nevertheless, the original sensor rotations are usually not delivered with the image data and therefore cannot be used for a geometric reconstruction of the image orientation. A quick change in viewing direction can lead to sensor vibrations, which can not always be captured by the provided 3rd-order rational polynomial coefficients of the satellite images. Vibrations of the pitch satellite angle (along track) influence the heights of reconstructed points, so it is possible to determine the effect of the vibrations by comparing a generated height model with a reference height model (Jacobsen 2018). As a result, satellite-based digital elevation models show systematic height errors (deformations, inclinations), due to insufficiently known image orientations determined with a typical small number of control points for RPC-bias correction.

Accurate attitude estimation of satellite platforms is the main requirement to provide a good geometric performance of remote-sensing imagery. The attitude is estimated by interpolating the measurement data provided by the star-trackers and gyros, which are sampled at a low frequency. However, when the sampling rate of attitude information is low, compared to the high frequency attitude jitter, the images processed by the geometric correction suffer from distortion or deformations (Teshima et al. 2007). Many high resolution satellites often experience attitude oscillations, causing image distortions and thus affecting the geo-positioning accuracy (Amberg et al. 2013; Tong, L. Li, et al. 2015; Ayoub et al. 2008; Mattson, Robinson, et al. 2010; H. Pan et al. 2013; Robertson 2003; Schwind et al. 2009; Toutin 2011). For example, authors of (Tong, L. Li, et al. 2015) detect and estimate the periodic distortions of the chinese Ziyuan-3 satellite with a frequency of 0.67 Hz, bringing improvements to the ground check points discrepancies at sub-pixel level. A method for correcting the Ziyuan-3 image distortions caused by satellite jitter is also proposed by Wang et al. (2016) showing a decrease of distortions across and along the track to less than 0.2 pixels as determined from known GCP positions. In (Schwind et al. 2009) oscillations of about one pixel in Advanced Land Observation Satellite (ALOS) prism images (2.5m/pixel) are identified. Ayoub et al. (2008) detected in Quick Bird images two distortions with frequences of 1 Hz and 4.3 Hz and amplitudes of 2.5 m and 0.1 m, respectively. As VHR satellite imagery continuously increase in usage, it is of high importance to develop methods for detecting and compensating image distortions caused by satellite attitude oscillations. In literature, there are two major approaches that address this research topic: (1) Methods based on parallel observations between two sensors, that acquire overlapping images

in a very short time interval (micro seconds) along the trajectory (Mattson, Boyd, et al. 2009; Iwasaki 2011; Mumtaz et al. 2013; Y.-h. Jiang, G. Zhang, Tang, D. Li, et al. 2014) and (2) Methods based on a rigorous model (Lehner et al. 2003; Schwind et al. 2009; R. Li et al. 2011; Bostelmann et al. 2011; Gwinner et al. 2010), but these are often replaced by the RFM. Nevertheless, the RPCs may contain the residuals of the approximation, in case the attitude oscillations occur in a higher order than the polynomial order of the coefficients in RPCs (Tong, L. Li, et al. 2015; Daniela Poli and Toutin 2012; Tong, L. Li, et al. 2015). These findings prove that image geometry should be corrected before the generation of RPCs. However, satellite image geometry is directly dependent on the accuracy of the attitude information and in situations when the attitude oscillations have a high frequency, they can not be detected by the attitude measuring sensors. Thus, the above-mentioned methods might not always work in practice. Therefore, we propose a new method based on vendor provided RPCs and a high resolution elevation model for improving the geometry of (Pléiades) VHR satellite images.

Literature on 3D reconstruction from stereo and tri-stereo satellite images

The geometric reconstruction from satellite image data poses special requirements for dense image matching algorithms. While existing algorithms have largely been optimized for the normal case (stereo image pair with parallel viewing direction) and the convergent case (many different image pairs cover the same scene), for satellite image data special challenges arise. On the one hand the corresponding sensor model must be supported (pushbroom as in the case of Pléiades and WorldView-3), on the other hand, the used dense matching algorithm must be adapted to the different viewing directions of the stereo-pairs. The built-up areas present an additional challenge to the matching algorithm because of the occlusion and shadowing effects in different viewing directions.

In recent years, the potential of stereo (tri-stereo) acquisition from high-resolution satellite images for surface modelling has become an interesting research topic that has been addressed in various publications. The radiometric and geometric characteristics of Pléiades imagery with focus on digital surface modelling are analyzed by Poli, Remondino, et al. (2013), where they use a "triple stereo" scene over Trento urban area in Italy. They describe street canyons as an issue, although these are visible in the 3D model. d'Angelo and Reinartz (2011) show that the reconstruction of bridges from WorldView-1 images (with same ground resolution as Pléiades) can be problematic. In contrast, Poli, Remondino, et al. (2013) use standard software to process the satellite images (like Match-T, Pix4DMapper), which as a black box solution, does not allow the full information to be used, but which can be used and is highly relevant for the detection of small objects (that are represented by only a few pixels in the images). Hüseyin Topan et al. (2013) state that Pléiades satellite images from "along track stereo" scenes are rarely used for 3D reconstruction, especially in urban areas. When reconstructing high, narrow street corridors between buildings, the problem often arises that the ground is only recognizable in one (or no) image and thus the reconstruction becomes impossible. In the case of airborne image sensors, this is avoided by the very high longitudinal and cross-overlaps between the images. Unlike aerial images, the lower pixel resolution of the satellite imagery, the occlusion problems (due to their wide base lines), and image matching

issues can notably influence the accuracy of the satellite-based DSM data. Different interpolation method were adopted for dealing with the void areas and imperfections in the satellite-DSM data (d'Angelo, Lehner, et al. 2008, Gharib Bafghi et al. 2016, Grohman et al. 2006), but still these are not completely satisfactory due to interpolation deficiencies such as blurring effects at buildings boundaries.

Although Pléiades system has clear advantages due to its agility, however a stereo configuration of the satellite images is still severely restricted to the acquisition geometry (Greslou et al. 2012). The advantages of using tri-stereo instead of stereo image pairs are described by Piermattei, Marty, Karel, et al. (2018) and Rupnik et al. (2018) where the nadir image increases the DSM completeness, reducing the occlusions usually caused by larger convergence angles on ground. Additionally, they investigate in detail the influence of the acquisition geometry (viewing and incidence angles) of VHR imagery on the DSM accuracy. The quality of the reconstructed elevation models from VHR satellite images was assessed in comparison with reference data such as ground control points and airborne lidar showing promisins results (Bernard et al. 2012, Perko, Raggam, Gutjahr, et al. 2014, Panagiotakis et al. 2018, Zhou et al. 2015, F. Hu et al. 2016, M. A. Aguilar, Nemmaoui, et al. 2019). The capacity of the Pléiades system in performing 3D mapping was analysed by Bernard et al. (2012), where they used 17 images acquired from different point of views over Melbourne, Australia. Their results showed that by means of "triple stereo" configurations reliable digital surface models can be generated in urban areas. Unfortunately, neither the used software used nor the applied matching algorithm are explained, which have a decisive effect on the results. Abduelmula et al. (2015) analyzes the potential of satellite images regarding a rapid evaluation of urban environments. They realize a comparison between 3D data extracted from tri-stereo and dual stereo satellite images, combined with pixel based matching and Wallis filter, in order to improve the accuracy of 3D models especially in urban areas. Their result showed that 3D models achieved by Pléiades tri-stereo outperformed the result obtained from a Geo-eye pair, both in terms of accuracy and detail. This could mean that tri-stereo images can be successfully used for urban change analyses. In Panagiotakis et al. (2018) the authors analyse and validate the potential of high-resolution DSMs produced from stereo and tri-stereo Pléiades-1B satellite imagery acquired over the Athens Metropolitan Area. From their tests, it resulted that the tri-stereo model shows the best performance in height accuracy when compared with elevations measured by a differential global positioning system. The potential of VHR optical sensors for 3D city model generation has been addressed in (Flamanc et al. 2005, Poli, Fabio Remondino, et al. 2015, Sirmacek et al. 2012), showing promising results when compared to a lidar elevation model, although highlighting some difficulties in the case of small individual house reconstruction. Besides 3D building and urban modelling, other important applications of the DEMs derived by using Pléiades stereo and tri-stereo acquisitions are dedicated to the estimation of lava flow volumes (Bagnardi et al. 2016) and changes in height produced by earthquakes (Zhou et al. 2015).

Much of the previous research using WorldView-3 satellite images is focused on their high resolution multi-spectral information, with applications in topographic mapping, land planning, land use, land cover classification, feature extraction and change detection, etc. (Alabi et al. 2016; Asadzadeh et al. 2016; Bostater Jr et al. 2017; Koenig et al. 2016). The 3D potential of WorldView-3 stereo images is addressed

by F. Hu et al. (2016), Mandanici et al. (2019), and M. A. Aguilar, Nemmaoui, et al. (2019), where they evaluate the accuracy of the reconstructed DEMs over different land covers: mountainous, bare soil, urban areas, and agricultural plastic greenhouses. A new algorithm for generating high quality digital surface models is proposed by Rupnik et al. (2018), where dense image matching method is applied for multi-view satellite images from Pléiades and WorldView-3. The achievable information content extraction from WorldView-4 stereo imagery (with same resolution as WorldView-3) by manual and automatic building and road extraction in an urban area is investigated by Sefercik et al. (2020).

To the best of our knowledge, only few previous scientific studies use high resolution stereo satellite images for 3D building model reconstruction (Arefi et al. 2013, Lafarge et al. 2006, Sirmacek et al. 2012, Duan et al. 2016). Partovi et al. (2019) proposes a new automatic multistage hybrid method using DSMs and orthorectified panchromatic and pansharpened data for an automatic 3D building model reconstruction from VHR satellite imagery. A quantitative and qualitative evaluation on 3D building models was presented in Moreira et al. (2013), where DSMs from a GeoEye-1 stereo-pair, from an aerial block and from lidar were employed. Their results show that the percentage of correctly reconstructed building models is very similar for airborne and lidar data, while for GeoEye data it is lower.

Knowledge of building height is a critical point in 3D city modelling for understanding the growth pattern of an urban development process. High resolution optical imagery can represent a source for detailed information within urban environments. Nevertheless, few studies have been conducted to estimate building height from high spatial resolution satellite imagery (Esch et al. 2020, Geiß et al. 2020, Mahendra et al. 2019). For example Y. Cao et al. (2021) adopt a multi-spectral, multi-view and multi-task deep neural network for estimating building heights from the high resolution ZY-3 satellite images at a spatial resolution of 2.5 m. Their height prediction results over 42 cities in China show accurqacies between 1.5 and 8.4 m, indicating a good potential for an automated large-scale building height mapping. Urban development analysis was investigated by Prakash et al. (2021) where Cartosat-1 stereo images at 2.5 m resolution over cities across India were used. The experimental results indicate the possibility of acquiring 50% of building heights within a +/-10 m error limit, showing the suitability of heights information extraction for macro-scale applications, but not for individual building modelling. Hence, besides a thorough investigation of the 3D capabilities of Pléiades and WorldView-3 sensors, special attention is dedicated to the single isolated objects like vehicles, trees and buildings that can be automatically reconstructed and detected in the photogammetrically derived elevation models, by estimating their heights as a function of object type and size.

Literature on point cloud classification

For a better understanding of the 3D scenes generated by satellite image matching, classification plays a significant role. In this context, this refers to the assignment of semantic labels to points, on a per-point basis (Otepka et al. 2013). Usually classification tasks are divided into two main categories: human- and machine-based classifications. The human-based classification is considered the most precise one, since the confluence of additional orthophotos, qualification and human experience lead to a performance that has not been exceeded by machines so far, especially

in terms of accuracy. On the other hand, human classification requires much more time when compared to computers, which are very fast. The machine-based classifications can be characterized either by a supervised or by an unsupervised learning. In supervised learning, a section of the data must be manually classified, hence an initial human classification input is required. These pre-classified data are then used for training the model and for supervision and should optimally contain the entire heterogeneity of the total area to be classified. Various methods of classification exist to optimally deal with this step, e.g. Regression trees (CARTs, Tan et al. 2005) or random forests (Breiman 2001). In unsupervised learning, no pre-classification is required, since the algorithm works on its own by simultaneously calculating the optimal number of classes, learning and applying the classification rules. The main advantage of supervised machine learning algorithms is that the resulting classes are fully in conformance with the human-defined ones, while the unsupervised algorithms return classes that might not be interpretable. However, both methods can be externally influenced and improved by integrating additional expert knowledge.

Point cloud classification, i.e. semantic segmentation of point clouds, has always been an essential and challenging task with applications in 3D city modelling, urban planning, mapping and monitoring, autonomous-driving, virtual/augmented reality, and robotics. An established and currently available classification method for the semantic segmentation of point clouds is represented by deep learning. As a dominating technique in Artificial Intelligence, deep learning has been successfully used for solving different 2D vision problems. However, typical deep learning methods developed for 2D images can not be directly applied to 3D point clouds. The increase of 3D data from different sensors together with the powerful computational resources such as GPUs have accelerated the fast-growing field of 3D deep learning. Even if it might face various processing challenges, deep learning on point clouds has become thriving, with numerous methods being proposed to address different problems in various applications from computer vision, robotics, and autonomous driving. Among the most relevant deep learning-based classification methods to be mentioned are: PointNet and PointNet++, Spherical CNN (Spherical Convolutional Neural Network), SGCNN (Sparse Generalized Convolution Neural Network), SO-Net (Self-Organizing Network), PointWeb, SpiderCNN, PoinwiseCNN, VoxNet (Guo et al. 2020). A comparison of deep learning methods for airborne lidar point clouds classification (PointNet++, SparseCNN, and KPConv) is investigated in (N. Li et al. 2021).

Many of the existing classification algorithms have been developed for laser scanning point clouds (for e.g. Gibbins et al. 2009, Mallet et al. 2011). Whereas these classification methods are mainly based on the point geometric features, in contrast, the classification of photogrammetric point clouds can be carried out using both geometric and radiometric information. While Z. Zhang et al. (2016) limit their classification approach to small ALS scenes, which do not include any color information, Rau et al. (2014) use RGB aerial photos acquired obliquely and vertically for defining a set of classification rules that are specific for urban environments. Niemeyer et al. (2014) integrate random forests into the conditional random field framework, but no color information is used. Especially, in lidar data, the separation between high vegetation and trees is based on the partial penetration of vegetation, information which is not available in the optical data from aerial or satellite images. Due to the existence of noise, occlusions and different objects types with various

sizes and shapes, the classification of 3D point clouds remains a challenging task.

Most of the previous studies on classification of 3D point clouds generated by image matching use rasterization or voxelisation, which reduces their full 3D content. For example, Gerke et al. (2013) converted the point clouds into a voxel representation and segmented them by adopting a "Random trees" machine learning technique and a supervised method (Markov-Random-Field). Finally, point clouds derived from image matching of airborne oblique images over two urban areas were classified into the following classes: façade, roof, rubble, sealed ground, and trees. In Modiri et al. (2015) the authors propose a region-growing technique to classify buildings and vegetation from stereo UltraCam-X matched point clouds, by using colour information and vegetation index. In contrast, in their approach Tran et al. (2018) analyse the results of two supervised classification algorithms on original 3D point clouds derived from high-resolution aerial images over an urban area with GSD of 6 cm.

Benchmarking datasets containing reference data are of great importance in evaluation tasks, being also used as training data in classification problems. Well known available benchmarks for 3D point cloud semantic segmentation are the following: the Oakland 3D point cloud dataset, which contains laser data collected from a moving platform in an urban environment (Munoz et al. 2009); the Sydney Urban Objects data set (De Deuge et al. 2013) and IQmulus and TerraMobilita Contest (Vallet et al. 2015) use mobile laser scanning for the 3D point cloud acquisition in dense urban environments; the ISPRS benchmark on urban object classification and 3D building reconstruction (Rottensteiner et al. 2012), and Semantic3D.Net: a new large-scale point cloud classification benchmark for natural and urban scenes (Hackel et al. 2017).

Nevertheless, in comparison with lidar and aerial images, the 3D point clouds generated by image matching from satellite images have a significantly lower quality. Hence, there is a small amount of work exploring the semantic segmentation of satellite image-derived point clouds. In their investigation, Leotta et al. (2019) develop an end-to-end system for segmenting buildings and bridges from terrain, by using point clouds derived from WorldView-3 multi-view satellite imagery. The reduced research in this direction might be caused by the low 3D quality of the obtained point clouds, due to the smoothing effects of the used dense image matching algorithm, as well as the 3D reconstruction difficulties encountered in the occluded, non-textured areas, water surfaces and repetitive patterns. An additional problem in deep learning is the requirement of extensive training data, which is not always available. Nevertheless, to investigate the quality of the photogrammetrically derived point clouds from Pléiades and WorldView-3 stereo/tri-stereo scenes, a supervised machine learning algorithm using decision trees is applied, for classifying the 3D point clouds into the following classes: ground, trees, roads, buildings, and vehicles. The quality of the adopted classifier was then assessed by analysing its repeatability and transferability performance.

In the European context, the use of high and very high-resolution satellite images was investigated for mapping tasks (EuroSDR project and research training "Mapping using high-resolution satellite imagery") and 3D information extraction (Daniela Poli and CARAVAGGI 2012). However, this has been done for manual evaluation or in the context of accuracy studies and covers only a small section of

the necessary wider investigations, like for instance high-frequency monitoring and automatic interpretation. The Copernicus services of the European Union and space agency, however, use Sentinel data, which with a coarse GSD of 10 m do not provide the required resolution for questions in the field of infrastructure. It is therefore commercial or national programs that provide high-resolution satellite imagery.

1.3 Research Questions

Nowadays, remote sensing scientists often begin a new project with the question: **Aerial or satellite imagery?** For long term projects, this question may arise many times, due to unforseen circumstances that might change the ability to collect data and the reliability of the data, since technology, budgets and government restrictions are continuously changing. Hence, a very careful look at the key differences between aerial and VHR satellite imagery will bring insight that better suits the needs for achieving optimal results in any remote sensing endeavour.

The derivation of DEMs from VHR stereo satellite imagery has become an interesting option and research topic, because the sub-meter GSD suggests that accuracy in the dm domain is possible. While previous research demonstrates the capacity of VHR stereo satellite imagery for DEM derivation, questions about the accuracy of the derived models and their full potential were not entirely investigated. Specifically, the focus of the thesis falls on the exploration of Pléiades and WorldView-3 stereo and tri-stereo scenes with the main purpose in finding the answers to the following research questions, which can be grouped into three main topics:

1. **Quality assessment and improvement of satellite elevation models**
   - What is the accuracy of elevation models derived from Pleiades and World-View-3 stereo/tri-stereo satellite imagery?
   - What is the impact of the different acquisition geometries (stereo / tri-stereo, ground sample distance, viewing and incidence angles) on DEMs derivation?
   - Which are the strengths and weak points of the satellite-based elevation models?
   - Are standard methods (RPC estimation, global transformations) sufficient to remove systematic errors or are more complex methods required for improving the accuracy of VHR satellite-based elevation models?
   - Which form of reference data is appropriate for removing systematic height errors in VHR satellite-based elevation models: ground control points, or terrain models of high or low resolution?
   - Can the distortions in image geometry be explained by satellite sensor vibrations?

To answer these questions and to explore the capabilities of VHR Pléiades and WorldView-3 satellite stereo/tri-stereo imagery for DSMs and Digital Terrain Models (DTMs) generation, four study sites with different land cover types were chosen: a rural, agricultural and forested area in Austria (Allentsteig), two urban areas: Barcelona and Stockholm, and the high Himalaya moutaineous range in Nepal, Asia. Due to their distinct topographic conditions (flat, hilly, mountaineous) and landscape type (rural, urban, grasslands, forests), the chosen study areas bring unique and particular challenges in the photogrammetric processing.

The absolute vertical accuracy of the photogrammetrically derived DEMs was evaluated against measured GCPs, lidar derived elevation models (where available, like in the cases of Allentsteig and Barcelona study sites), and global ALOS 30 m elevation model (for Stockholm and Himalaya datasets). When using the lidar DTM for the evaluation, the focus of the assessment analyses falls on the free, open and smooth areas (without any vegetation or artificial structures like buildings or infrastructure), where usually the DSM can be considered as DTM. Additionally, the geometric quality of the generated DEMs was analyzed with respect to the impact of the different acquisition geometries (stereo / tri-stereo, ground sample distance, viewing and incidence angles) on the estimated surface and its properties. For the analysis and estimation of VHR satellite image distortions caused by attitude oscillations in flight direction, a new method was proposed. Its effectiveness is then proved by the improvement of the satellite-based DSMs geoposition accuracy.

2. Classification of 3D point clouds derived from VHR satellite stereo/tri-stereo scenes

- How repeatable is the process from acquisition to classification result?
- What is the impact of acquisition geometry and GSD on the classification result?
- What is the transferability and cross-time performance of the classifier? How does the classifier trained on a dataset perform on a new dataset for the same area, but from a different acquisition time?

In this case, the main aim was to understand the quality of the satellite image derived point clouds for semantic segmentation. Therefore, a supervised machine learning algorithm using decision trees was employed for classifying the tri-stereo satellite image-derived 3D point clouds into five different classes: ground, trees, roads, buildings, and vehicles. Moreover, through repeatability and transferability analyses the performance of the adopted classifier was assessed. Indicative answers to the classification-related research questions were given by using three tri-stereo satellite image acquisitions over the same area (Allentsteig).

3. Orthorectification of VHR satellite imagery

- What are the advantages and drawbacks of true orthophotos in urban areas?
- Which solutions could be adopted for enhancing the orthorectified satellite imagery in urban areas?
- What are the main differences between true, DTM- and constant height plane-based orthophotos?

Specific attention was dedicated to the improvement of orthophoto geolocation accuracy and visually appearance, principally in urban areas. For this, the Barcelona city was selected as investigation site and the satellite image geometry was corrected by adopting different elevation models: DSM (true orthophoto), DTM (DTM-based orthophoto) and a constant height plane (constant height plane-based orthophoto).

4. Potential of satellite products for 3D reconstruction of isolated objects

- Under which conditions image matching from satellite imagery works well and where not?

- What level of detail can be extracted from VHR satellite image products? What and which specific objects can be reconstructed or not? Is it possible to detect vehicles in the satellite data?
- Which geometric signature, conditions and minimum size must individual objects have, in order to be detected in the DSM, based on their reconstructed heights?
- What are the influence of different acquisition times, geometries and GSD on dense image matching quality for single objects?

As a key component in 3D reconstruction from imagery data sets, dense image matching process encounters difficulties in areas with no texture or similar patterns, at water-covered areas, and at very small objects. The later was subject to further investigations in the current work, since no research was conducted so far in this direction and in the context of VHR satellite imagery. Therefore, to assess the capability and limitative aspects of space-borne very high resolution imagery, specifically Pléiades (0.70 m) and WorldView-3 (0.31 m) imagery with respect to the reconstruction of small, isolated objects, three different object types were thoroughly analysed: parked vehicles, trees and buildings.

5. Visibility analysis from satellite imagery
- Which factors affect satellite image visibility?
- Which is the relation between satellite image acquisition parameters and visibility in image space?
- How could occluded areas mathematically be described based on satellite viewing geometry acquisition angles?

For this problematic, the stereo WorldView-3 satellite images acquired in July 2017 over Barcelona city were used. Depending on city geometry and satellite image acquisition parameters, the geometric and mathematical description of street visibility was thoroughly investigated.

1.4 Contributions

Overall, this dissertation investigates and explores the 3D capabilities of stereo, tri-stereo VHR satellite imagery. In contrast to the classic two-dimensional interpretation of satellite images, modern three-dimensional methods were used and, in particular, improved in order to be able to provide semantic 3D position information in an automated and timely manner. Given the above enumerated research questions and in order to reach the proposed objectives, the main contributions of this study can be grouped into two categories, as following:

(1) contributions developing new methodologies for improving the actual state of the achievable information from satellite imagery (in sections 4.3 and 4.4)

(2) contributions needed to better understand the actual state of available information from satellite imagery (in sections 4.1, 4.2, 4.5, 4.6, and 4.7).

(1). Contributions that develop new methods for improving the actual state of the achievable information from satellite imagery:

1. Developing and applying a method for the geometry correction of satellite imagery by using a high resolution elevation model

2. Analysis of satellite oscillation movements during data acquisition based on image corrections

3. Improvement of satellite-based DSMs

4. Improvement of orthophoto geolocation accuracy and quality, by adopting different elevation models: DSM (true orthophoto), DTM (DTM-based orthophoto) and a constant height plane.

(2). Contributions needed to better understand the actual state of available information from satellite imagery:

5. Description of a complete, improved, and automatic photogrammetric approach for satellite imagery processing chain

6. Analysis of systematic errors based on TPs residuals in image space

7. A comprehensive overview of the vertical accuracy of VHR satellite-based elevation models

8. New insights about the capabilities of VHR optical sensors with respect to satellite image orientation, dense matching and computation of digital elevation models, by comparisons between different data sources and acquisitions over same areas

9. Detailed analysis concerning the classification of 3D point clouds extracted from VHR satellite images with dense image matching

10. New aspects regarding the satellite-based point cloud classification based on machine learning method using regression trees concerning repeatability and transferability, which give an enhanced insight of the classification performance

11. First analysis on the potential of VHR tri-stereo satellite imagery for 3D reconstruction of small, isolated object regarding the achievable level of detail based on object geometric signature

12. Visibility analysis in satellite imagery for urban areas from a geometrical perspective of satellite position along its orbit with different in-track and cross-track viewing angles.

1.5 Publications

The following publications and scientific contributions resulted from the work presented in this thesis:

- **Papers published in peer-reviewed journals:**

    1. Loghin, Otepka-Schremmer, Ressl, et al. 2022: Loghin, A.-M.; Otepka-Schremmer, J.; Ressl, C.; Pfeifer, N., 2022. *Improvement of VHR Satellite Image Geometry with High Resolution Elevation Models.* Remote Sensing, 14, 2303.

    2. Loghin, Otepka-Schremmer, and Pfeifer 2020: Loghin, A. M., Otepka-Schremmer, J., Pfeifer, N. 2020. *Potential of Pléiades and WorldView-3 Tri-Stereo DSMs to Represent Heights of Small Isolated Objects.* Sensors, 20(9), 2695.

    3. Bachhofner et al. 2020: Bachhofner, S., Loghin, A. M., Otepka, J., Pfeifer, N., Hornacek, M., Siposova, A., Schmidinger, N., Hornik, K., Schiller, N., Kähler, O., Hochreiter, R., 2020. *Generalized Sparse Convolutional Neural*

Networks for Semantic Segmentation of Point Clouds Derived from Tri-Stereo Satellite Imagery. Remote Sensing, 12(8), 1289.

4. Zambanini et al. 2020: Zambanini, S., Loghin, A. M., Pfeifer, N., Soley, E. M., Sablatnig, R., 2020. *Detection of parking cars in stereo satellite images.* Remote Sensing, 12(13), 2170.

- **Papers published in conference proceedings:**

5. Loghin, Otepka-Schremmer, Karel, et al. 2019: Loghin, A. M., Otepka-Schremmer, J., Wilfiried, K., Pöchtrager, M., Pfeifer, N., 2019. *Accuracy Analysis of Digital Elevation Models from very High Resolution Satellite Imagery.* Proceedings of Dreiländertagung der DGPF, der OVG und der SGPF - Publikationen der DGPF, Band 28, 123-137 (including oral presentation).

6. Loghin, Pfeifer, et al. 2020: Loghin, A. M., Pfeifer, N., Otepka-Schremmer, J., 2020. *Supervised Classification and its Repeatability for Point Clouds from Dense VHR Tri-Stereo Satellite Image Matching Using Machine Learning.* IS-PRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2, 525-532. (peer-reviewed conference proceedings paper, including oral presentation)

- **Other related publications:**

7. Loghin, Oniga, et al. 2018: Loghin, A. M., Oniga, V. E., Giurma-Handley, C., 2018. *3D Point Cloud Classification of Natural Environments Using Airborne Laser Scanning Data.* American Journal of Engineering Research (AJER), 7(2), 191-197.

8. Otepka-Schremmer et al. 2019: Otepka, J., Pöchtrager, M., Hollaus, M., Loghin, A. M., Piermattei, L., Briese, C., Pfeifer, N., 2019. *Landesweite Punktwolkenprozessierung mit OPALS am Supercomputer VSC3.* 20. Internationale Geodätische Woche Obergurgl 2019, ISBN: 9783879076598, 165 - 176. (conference proceedings paper)

The work in this dissertation has been conducted as part of the following two research projects:

Note on authorship It is noted that sections 4.2.4, 4.3, 4.5, and 4.6 are largely taken from the following papers: Loghin, Otepka-Schremmer, Karel, et al. 2019, Loghin, Otepka-Schremmer, Ressl, et al. 2022, Loghin, Pfeifer, et al. 2020, and Loghin, Otepka-Schremmer, and Pfeifer 2020.

1.6 Thesis structure

The present thesis is organized as follows:

Chapter 1 - Introduction brings out the motivation and objectives for the research together with the current state-of-art for VHR satellite imagery and principal addressed research questions.

Chapter 2 - Methodological framework focuses on the photogrammetric processing chain for stereo/tri-stereo satellite images comprising image orientation based on the RFM improved with ground control points, dense image matching and 3D information extraction followed by DSM derivation. Additionally, the methodology for the supervised semantic segmentation applied to the satellite-based 3D point clouds is described here. This chapter can be considered as a theoretical baseline or as starting point for the further analyses, since it contains the detailed description of the photogrammetric workflow for the satellite imagery, without having any new methodological contributions.

Chapter 3 - Materials presents the four study areas of interest: Allentsteig, Barcelona and Stockholm cities, and a region in the Himalaya mountains, with their particularities regarding the landscape type (rural, urban, grassland, forest), topographic conditions (flat, hilly, mountainous terrain), and country location (Austria, Spain, Sweden, and Nepal). The stereo/tri-stereo VHR imagery from Pléiades and WorldView-3 satellite sensors for the areas of interest are described in section 3.2.

Chapter 4 - Experimental analyses and results contains the results for satellite image orientation, DSM generation from satellite imagery, improvement of satellite-based DSM, image orthorectification, supervised point cloud classification, height of small isolated objects estimation and visibility analysis, followed by their quality assessment and discussion. This chapter contains the two types of new contributions mentioned in Section 1.4.

Chapter 5 - Summary and conclusions outlines the conclusions with regard to the potential, capabilities and limitations of VHR satellite imagery in the context of 3D reconstruction and classification applications.

<div align="right">

2

</div>

Methodological framework

2.1 End-to-End photogrammetric workflow for VHR satellite images

Numerous applications in remote sensing require high and accurate up-to-date mapping products in form of DSMs, DTMs or their differences, which are the normalized Digital Surface Models (nDSMs). All this information can be extracted from stereo/tri-stereo/multi-view satellite imagery through photogrammetric techniques. A complete workflow for 3D information extraction from optical satellite images is shown in Figure 2.1. The input data consists of (1) multi-view satellite images (one or multiple stereo/tri-stereo pairs) with their related metadata, which are the sensor model for each image in the form of RPCs and (2) reference GCPs with their 2D image coordinates, where the image coordinates are usually manually measured within an image-based interactive measurement tool. This information is then employed in an automatic workflow as available in the Trimble Inpho software environment. For extracting 3D information, the following main steps were applied individually to each satellite image data set:

1. import of image with metadata information (RPCs) and GCPs;
2. GCP measurement in image space;
3. orientation refinement based on GCPs and TPs;
4. dense image matching for 3D reconstruction.

The workflow as applied in this study starts with satellite image triangulation and dense image matching, followed by 3D reconstruction, DEM interpolation, geolocation improvement and accuracy assessment (Figure 2.1).

The final output of the process are the 3D point clouds and corresponding DSMs, whose geolocation is further assessed against reference DTM and improved by means of the Least Squares Matching (LSM) technique.

2.1.1 Sensor modelling

The two most widely used imaging geometry models in photogrammetry and remote sensing are the physical and generalized sensor models (Y. Hu et al. 2004). Physical models, also named rigorous, parametric, or deterministic represent the physical imaging process, by fully reflecting the sensor viewing geometry (M. A. Aguilar, F. J. Aguilar, Agüera, et al. 2007). Normally uncorrelated because of their physical significance, the involved parameters describe the position and orientation of a sensor with respect to an object-space coordinate system. Based on the collinearity condition, physical models are rigorous and suitable for adjustment by analytical triangulation (Tao and Y. Hu 2001). This model yields in high accuracies (typically
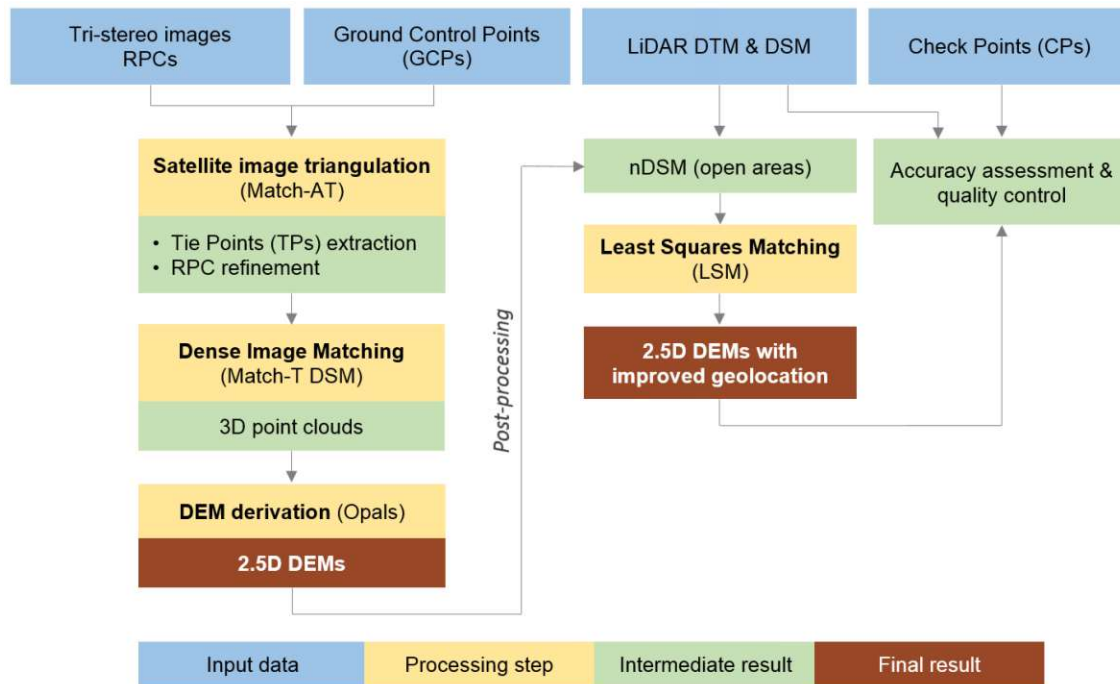
Figure 2.1: Photogrammetric workflow for 3D information extraction from stereo/tri-stereo satellite imagery.

a fraction of one pixel), but is complex and varies depending on different sensors types (e.g. Pushbroom, Frame, Whiskbroom, Panoramic, etc.). Due to the increasing availability of imaging systems, implementing of a physical model for each individual sensor is not a feasible solution from an application point of view. Moreover, information about the camera model, ephemerides and satellite attitude may not be available to the end users, since they are kept confidential by some commercial satellite image providers (Tong, S. Liu, et al. 2010). Therefore, developing a mathematically complex rigorous sensor model without knowing its imaging parameters (e.g. imaging geometry, relief displacement, Earth curvature, atmospheric refraction, lens distortion, etc.) is very challenging, expensive, time-consuming, and error-prone.

The RFM is a generalized sensor model used in photogrammetric mapping as an alternative solution for replacing the rigorous physical sensor models (Gene Dial et al. 2002). This became a standardised way for economical and fast mapping from high-resolution images and has been in practice for more than two decades. The RFM containing eighty RPCs provides a direct mapping from 3D object space coordinates to 2D image coordinates (Daniela Poli and Toutin 2012). The projection from object to image space is denoted by $RPC : R^3 \rightarrow R^2, (X, Y, Z) \rightarrow (r, c)$. Hence, in the RFM, image pixel coordinates (r, c) are expressed as ratios of polynomials of ground coordinates (X, Y, Z), as follows:

$$
\begin{aligned}
r_n &= \frac{P1(X_n, Y_n, Z_n)}{P2(X_n, Y_n, Z_n)} \\
c_n &= \frac{P3(X_n, Y_n, Z_n)}{P4(X_n, Y_n, Z_n)}
\end{aligned}
\tag{2.1}
$$

where (r_n, c_n) are the normalized row and column index of pixels in image, (X_n, Y_n, Z_n) are the normalized object point coordinates on the ground and P_i are the third-order polynomial functions in the following form:

$$
\begin{aligned}
P = \sum_{i=0}^{m1}\sum_{j=0}^{m2}\sum_{k=0}^{m3} a_{ijk}X^i Y^j Z^k = \\
= a_1 + a_2 Z + a_3 Y + a_4 X + a_5 ZY + a_6 ZX + a_7 YX + a_8 Z^2 + a_9 Y^2 + \\
+ a_{10}X^2 + a_{11}ZYX + a_{12}Z^2 Y + a_{13}Z^2 X + a_{14}Y^2 Z + a_{15}Y^2 X + \\
+ a_{16}ZX^2 + a_{17}YX^2 + a_{18}Z^3 + a_{19}Y^3 + a_{20}X^3
\end{aligned}
\tag{2.2}
$$

where a_{ijk} are the polynomial function coefficients called RPCs (Tao and Y. Hu 2001).

2.1.2 RFM bias compensation

Rational functions have proven to be a feasible substitute model to many physical sensor models due to their high approximating accuracy in geo-positioning. However, because of various errors in the satellite ephemeris and instrument calibration, the RPCs provided by the satellite image vendors may not always represent the real imaging process with high accuracy. Therefore, there is a clear need to correct the systematic biases in order to meet the requirements of high-precision topographic mapping (Shen, B. Liu, et al. 2017). The RPCs can be refined in the image or ground space by using additional ground control information. High sensor orientation accuracy to sub-pixel level can be achieved by performing a bias-compensation bundle adjustment with one or more ground control points (Grodecki et al. 2003, Y. Hu et al. 2004, Fraser and Hanley 2005). By integrating additional parameters for image shift and drift, the basic model of Equation 2.1 takes the following form:

$$
\begin{aligned}
r_n + a_0 + a_1 r + a_2 c = \frac{P1(X_n, Y_n, Z_n)}{P2(X_n, Y_n, Z_n)} \\
c_n + b_0 + b_1 r + b_2 c = \frac{P3(X_n, Y_n, Z_n)}{P4(X_n, Y_n, Z_n)}
\end{aligned}
\tag{2.3}
$$

From the above formulation there are three different options for the bias-compensation model:

1. Shift model defined by two parameters: a_0 and b_0
2. Shift and drift model defined by four parameters: a_0, b_0, a_1, and b_1
3. Affine transformation model defined by six parameters: a_0, b_0, a_1, b_1, a_2, and b_2

In the first case, the shift parameters will perform an image coordinate translation where a minimum of one GCP is required. Precision could be enhanced by more GCPs, but their number and location is not important. On the other hand, in cases (2) and (3) the situation is slightly different, since a non-conformal transformation of the image coordinates is taking place and the absolute orientation process implies a modification of the initial orientation given by Equation 2.1. Here both position and number of implied GCPs plays a significant role, with a practical minimum number being 4-6 (Fraser and Hanley 2005).

Both Pléiades and WorldView-3 are pushbroom scanners, which use a linear array of detectors known as Charged coupled devices (CCD) arranged perpendicular to the flying direction. This allows the pushbroom scanner to record one line of an image simultaneously (line perpendicular to the flying direction), as the platform moves forward. One drawback of this type of sensor is that the detectors in the array can have varying sensitivity and if they are not perfectly calibrated, this can result in stripes in the recorded data. The generalized sensor model used for image orientation is based on the eighty RPCs, available as metadata information for each image. The reported geo-location accuracy for the Pléiades RPCs is of 8.5 m CE90 (circular error at 90% confidence) corresponding to the nadir view (Astrium 2012), whereas for the WorldView-3 is of 3.1 m CE90 (Comp et al. 2015).

For obtaining a sub-meter accuracy, the RPCs are refined by including GCPs in the orientation workflow (Heipke 1997, Toutin 2004). GCPs should be distributed homogeneously over the entire satellite scene to enable estimation of correction models which apply to the area imaged. As described in Section 3.3, the target positions of these points correspond to fixed details on the ground, such as road marks (e.g. pedestrian crossing lines), corners of field, parking lots and pavement boundaries, road surface changes and intersections. GPCs can be measured with accuracy of a few centimeters, e.g., by GNSS RTK. However, considering the accuracy of standard orthophotos (0.10 m in Austria) and the vertical accuracy of the DTM (0.12 m) also the measurement of control points is possible without going into the field. During image orientation, TPs are automatically extracted using a multi-ray image matching method.

The main steps for the satellite image triangulation used in this thesis are described as follows:

1. **GCPs measurement in images.** The number of GCPs is different for each image dataset, because of the different extend and visibility in the scenes. The GCPs were manually measured in each image of the Pléiades or WorldView-3 tri-stereo pairs using the multi-aerial viewer tool of the Match-AT photogrammetric software, which allows a simultaneous display of the images. This step is performed in order to stabilize the image block and to achieve higher accuracies, by improving the image orientation provided through the given RPCs' values.

2. **TPs extraction and RPCs refinement.** The automatically extraction of interest points is performed by using a feature detector algorithm. Based on the description of the point neighbourhood, performed by a feature extractor, points with a poor description are eliminated, and finally matches between images are determined (Trimble 2016a). The TPs were further used together with the GCPs in a bundle block adjustment, to refine the initial values of the RPCs. During the adjustment, a maximum number of six parameters (affine transformation in image space) can be computed: two offsets, two drift values and two shear values (for each image). Depending on their significance, only a subset of these corrections could be computed by the software, namely: two shifts (on r and c) and a scale on r. Simultaneously, TPs are automatically extracted by applying Feature Based Matching (FBM) using the Förstner operator and refining them with LSM subsequently (Trimble 2016b). From the resulted statistics, TPs with residuals (in image space) larger than one pixel were considered as blunders and therefore, rejected. The RPCs were

then refined through a subsequent adjustment procedure where the differences between image- and backprojected- (with the RFM) coordinates of the GCPs and TPs are minimized.

3. **Geo-positioning accuracy analysis.** To evaluate the numerical accuracy of the georeferenced images, the GCPs and Check Points (CPs) were used. They were manually acquired from the available orthophoto and lidar DTM as described in Section 3.3. To verify the georeferencing, the RMSE values in latitude, longitude and elevation between the adjusted coordinates (computed by the software) and the reference values are provided.

For the current work, the entire photogrammetric workflow was implemented in the Inpho 8.0 software from Trimble, designed to perform precise image block triangulation through bundle block adjustment and 3D point cloud reconstruction using dense image matching techniques for push-broom cameras. The same processing chain was followed for each Pléiades and WorldView-3 stereo/tri-stereo scenes corresponding to the four test areas.

2.1.3 Dense image matching and 3D information extraction

With a history of more than fifty years, image matching is one of the key factors in 3D modelling and mapping, used to identify homologous objects or pixels within oriented images and to derive 3D information (point clouds). Defined as the establishment of correspondences between primitives extracted from two or more images (Schenk et al. 2000), image matching determines the 3D coordinates of matched feature points via a collinearity or projective model. In image space a depth map (containing depth values in each image pixel) is generated, while the corresponding outcome in object space is the 3D point cloud. For a stereo image pair, the automatic measurement of disparities (parallax - horizontal discrepancy) by establishing dense and accurate image correspondences remains a challenging task (Remondino et al. 2014).

Currently available image matching algorithms can be divided into area-based (Förstner 1982), feature-based (Förstner 1986), and pixel-based matching (Hirschmuller 2007). Based on the involved number of images (stereo-pairs or multiple images) the algorithms can be grouped into stereo-matching (Hirschmuller 2007, Haala et al. 2012) and multi-view matching (L. Zhang 2005, Pierrot-Deseilligny et al. 2006). Alternatively, image-matching approaches are often classified into sparse and dense matching or according to (Szeliski 2010) into local or global.

Traditional stereo-matching typically uses feature based algorithms, which firstly extract the feature points and secondly, search the corresponding features in the overlapping images (Heipke 1992). Due to the significant advances in dense matching over the last few years there have been developed various free, open-source and commercial software programs, which have implemented different matching strategies (VisualSFM, CMVS, Pix4D, Agisoft, Sure, etc.). While all software products can deal with central-perspective images, satellite data (push-broom sensors) require the implementation of the corresponding sensor model or the division of the scene into small image sections, where central perspective images can be simulated. Various software packages such as, for example, ENVI (Exelis), ERDAS (Hexagon Geospatial), Geomatics (PCI Geomatics), Socet SET (BAE Systems), ArcGIS (ESRI), Match-T (Trimble) are widely used to extract DSMs from very

high resolution stereo/tri-stereo satellite imagery.

By their overlapping characteristic, the available stereo/tri-stereo Pléiades and World View-3 VHR satellite images for the four test areas, allow the extraction of 3D information by applying photogrammetric techniques. The corresponding pixels between images collected from the three different viewing directions (forward, nadir and backward) are found through dense image matching. For each study site, the workflow for 3D point cloud extraction from satellite images was performed in the Trimble Inpho photogrammetric software and is depicted in Figure 2.1.

The automatic image matching and DSM reconstruction processes were individually performed for each tri-stereo-scene by using the specialized module of the Inpho software, called Match-T DSM. Here, the DSM derivation is based on three matching strategies (Trimble 2016b):

(a) Least Squares Matching (LSM)

(b) Feature Based Matching (FBM)

(c) Cost Based Matching (CBM)

The ground coordinates of the corresponding image pixels (retrieved from image matching) are computed by applying forward spatial intersections by means of a least squares approach. This results in a "cloud" containing 3D points regularly distributed on the ground surface. Like in most image matching procedures, where image pyramids are used to reduce computation time (Heipke 1992), in this case, the iterative processing chain contains a number of ten pyramid levels. On each pyramid level three processes are performed: the matching of homologous image points, 3D intersection in object space and DEM modelling. For the first seven pyramids FBM (Förstner 1986) was used and the last three image pyramid levels were processed with Cost Based Matching (CBM). CBM is a pixel-by-pixel matching technique similar to the semi-global matching algorithm (Hirschmuller 2007, Förstner 1986, Heipke 1992, Wenzel et al. 2013). The CBM strategy within the Match-T DSM module uses a search-path in a so-called 3D-cost-cube for finding the corresponding pixels in images. The cost functions (e.g. correlation coefficient) are used to find the minimum cost path and each direction represents an X-Y movement in the image to match. Finding the pixel with the lowest cost generates a lowest-cost 3D model – surface model (Tong, S. Liu, et al. 2010). For each image pixel, the 3D object point coordinates are calculated by applying forward intersections, finally resulting in dense photogrammetric point clouds with three-dimensional coordinates in object space (i.e. X, Y, Z) and reflectance information from the three spectral bands (Red, Green and Blue).

Although dense image matching is a key component in photogrammetric 3D reconstruction and mapping, it encounters difficulties for water surfaces, areas with no texture or with a repetitive pattern appearance in the images, as well as for very small objects.

2.2 DSM derivation and accuracy assessment

For each study site, regular raster models of height values (DSMs) at high resolution (1 m for Pléiades and 0.5 m for WorldView-3) are generated by using a robust moving planes interpolation method. For each grid node all points within a circular neighbourhood (search radius of 1 m for Pléiades and of 0.5 m for WorldView-3)

are used to robustly estimate a best fitting tilted plane (minimizing the vertical distances); points classified as outliers are not considered in the plane fitting procedure. According to the input point density, this approach was chosen to be the optimal in terms of surface detail preservation.

The vertical accuracy assessment of the reconstructed DSMs was determined against the available elevations of the GCPs, CPs and reference lidar DTM in open areas. Hence, the vertical accuracy was determined by: (1) computing the Root Mean Square Errors (RMSEs) between reference and extracted elevations from each generated DSM for the GCPs and CPs; (2) visual analysing the difference elevation models (color-coded by elevations with corresponding color palette) and (3) analysing the histogram of difference values that should be around zero. For this, the distribution histogram was investigated by deriving statistic parameters such as mean, median, standard deviation, robust standard deviation (σ_{MAD}) and also RMSE. This was possible for the three Allentsteig and Barcelona study areas, where GCPs and reference lidar DTM are available. In the case of Stockholm and Himalaya datasets, the ALOS elevation model was used as reference. Here, the vertical accuracy of the reconstructed DSMs was evaluated for the entire areas by computing the difference elevation models to the ALOS DSM.

2.3 Satellite DSMs Improvement

As described above, after orientation a DSM can be provided. Such a DSM will typically contain 3 types of errors: (1) random (height) errors which are the results of random matching errors, (2) systematic height errors, which differ from the random errors by showing correlation (i.e. neighboring points have the same error), and (3) gross errors, which result from gross matching errors and are, to a very large extent, already eliminated by the robust measures taken in image orientation and DSM computation.

The systematic errors may have different reasons, and there are different strategies to measure their size and to compensate them.

2.3.1 Impact of number and distribution of GCPs on DSM accuracy

As the image georeferencing phase is of primary importance for subsequent processing, the satellite scenes were oriented through the rational function model based on the eighty polynomial coefficients provided by the supplier. However, as recommended, the geometry needs an improvement (Heipke 1997; Toutin 2004) and in order to reach a higher sensor orientation accuracy (to sub-pixel level), the model was further refined within a bias-compensation procedure by using GCPs and automatically extracted TPs, well distributed in the study area. To assure a sufficiently good satellite image orientation and to check if systematic errors can be compensated by residuals in image space, we used a higher number of GCPs for refining the provided RPCs, with no automatic extraction of TPs. The considered investigation makes use of a high number of GCPs homogenously distributed over the Allentsteig scene (Figure 2.2), which is a Pleiades triple stereo scene. Their positions were defined based on the reference orthophoto and the corresponding heights extracted

from a reference Lidar DTM. The 2D image coordinates of the GCPs were manually measured within the multi-aerial photo measurement view mode of the Match-AT satellite triangulation module from Sat-Master photogrammetric software, which allows a simultaneous display of the images. To evaluate the accuracy of the geo-referenced images, the RMSE between the adjusted coordinates (provided by the software) and the originally measured coordinates were computed, resulting in an overall RMSE of the residuals well below one pixel.
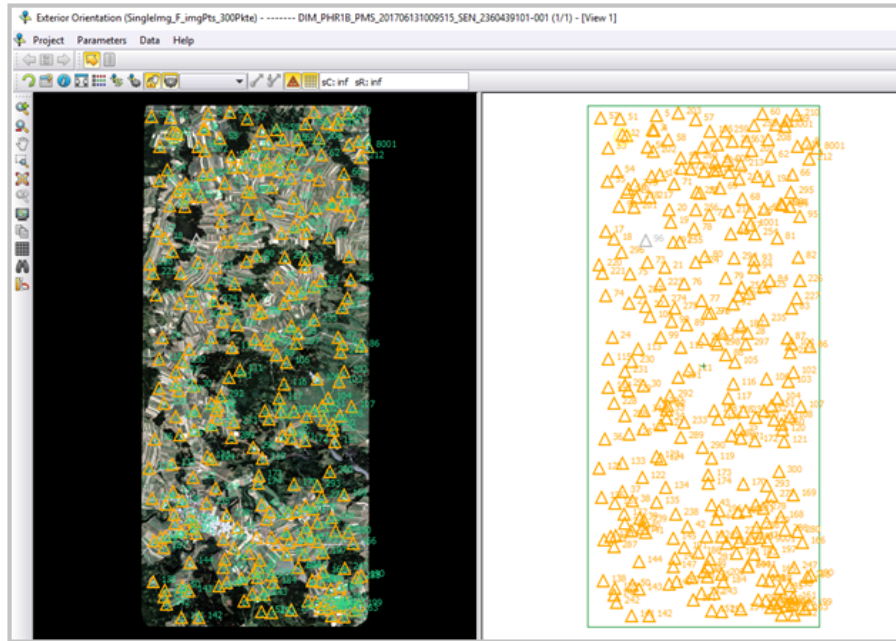

Figure 2.2: Overview of the manually measured GCPs distribution in Match-AT software.

Dense image matching was performed for each stereo pair i.e., forward-backward (FB), forward-nadir (FN) and nadir-backward (NB) and in a last step, regular raster models of height values (DSM) at 1 m resolution are generated by using a robust moving planes interpolation. A search radius of 1 m and a maximum of 10 nearest neighbours are used for the interpolation. According to the input point density, this approach was chosen to be the optimal in terms of surface detail preservation.

The accuracy (in planimetry and elevation) of the computed satellite image triangulation was evaluated by comparing the GCPs coordinates (measured either on the ground or on orthophoto and lidar) with those monoscopically measured on the oriented satellite images in a multi-aerial view-mode. The summary of the resulted statistics for the coordinates are then visualised with boxplots.

The vertical accuracy assessment of the reconstructed DSMs was determined against the available elevations of the GCPs and reference Lidar DTM in open areas by: (1) computing the RMSE between reference and extracted elevations from each generated DSM for the GCPs; (2) visual analysing the difference elevation models (color-coded by elevations with corresponding color palette) and (3) analysing the distribution histogram of difference values, by deriving statistic parameters such as mean, median, standard deviation, robust standard deviation (σ_{MAD}) and RMSE. After the quality checking, systematic errors were visible between the reconstructed DEMs and the LiDAR DTM. Therefore, an improvement strategy is required.

Obtained through photogrammetric processing, the satellite-based digital elevation models can be further improved by adopting several techniques. These can be either applied in object space (like the LSM technique) or in image space (through the proposed methodology of satellite image geometry correction).

2.3.2 Least Squares Matching technique

As a traditional photogrammetric technique used for establishing correspondences between images collected from different viewing points (or between consecutive images of a sequence), LSM determines the parameters of an affine transformation between corresponding patches of two or more images. Formulated for two-dimensional greyscale images (Förstner 1984), the technique can be applied to surface models with a grid structure, as well (Maas 2000, Ressl, Kager, et al. 2008). In this case, LSM acts very similar, by estimating an affine 3D transformation to align overlapping elevation models. The main idea is to find the corresponding location of a surface patch inside a window of one dataset in another dataset. This is usually done by estimating a transformation in such a way that the differences between the two input models inside the patch are minimized. In contrast to images, where the defined LSM patch size should be limited to small windows (due to different viewing directions and perspective image projection), LSM can be applied to satellite- or ALS-derived elevation models over large areas (due to the orthogonal projection of the models).

In these investigations, the LSM technique is used in two different ways:

1. **Global LSM** - for global improvement of the georeferencing of photogrammetrically derived DSMs

2. **Continuous LSM** - for modelling the periodic systematic elevation errors (waves) in object space (described in Section 2.3.3).

Applied to the photogrammetrically derived DSMs, the global LSM technique estimates an affine transformation between each input elevation model and the reference Lidar DTM over common open surfaces. For a reliable transformation estimation, some specific areas were not included during LSM, because of vegetation cover, matching errors and/or occlusion. Hence, from the input surface models, areas containing vegetation, rivers, lakes, buildings and other artificial objects on the ground were masked out and therefore not included in the computation. For this purpose, an above-ground mask was computed by using the volume-based algorithm developed by Piltz et al. (2016), having as input the photogrammetric sattelite DSM. Therefore, only smooth, object-free stable surfaces (called **open areas** in the following) were considered for the computation. Here, the differences in height between DSM and DTM are expected to be as low as possible (below 10 cm). In the case of Allentsteig study site, the percentage values of stable areas within each dataset are of approximately 37.8 % and of 32.8 % for Pléiades and WorldView-3 scenes, respectively. Subsequently, the resulted LSM parameters were used to transform each Pléiades and WorldView-3 point clouds, which were then reinterpolated into new elevation models.

The LSM affine 3D transformation with 12 parameters can be expressed in the following form:

$$p' = A(p - p_0) + t + p_0 \qquad (2.4)$$

where $p = (x \ y \ z)^T$ is the point given in the system of the dataset to be transformed (DSM), $p' = (x' \ y' \ z')^T$ is the same point given in the system of the fixed reference dataset (lidar DTM), $p_0 = (x_0 \ y_0 \ z_0)^T$ is the reduction point, $t = (d \ h \ l)^T$ the translation vector and the rotation matrix given by:

$$A = \begin{bmatrix} 1+a & b & c \\ e & 1+f & g \\ i & j & 1+k \end{bmatrix} \qquad (2.5)$$

Performed for open areas, the global LSM technique estimates a single global transformation parameter set, which is subsequently used to transform the Pléiades DSMs to the Lidar DTM. In contrast to global LSM where large open areas are considered for the computation, the continuous LSM technique uses smaller patches arranged along flight direction for modelling the periodic systematic elevation errors. Both LSM modes, global and continuous, were conducted in the scientific software OPALS (Orientation and Processing of Airborne Laser Scanning Data) (Pfeifer et al. 2014).

2.3.3 Image geometry correction

The proposed method for computing the satellite image geometry corrections employs a higher resolution elevation model (like in this case a Lidar DTM). The fully integrated methodology is depicted in Figure 2.3 and comprises three main steps:

1. Photogrammetric processing
2. Elevation difference computation
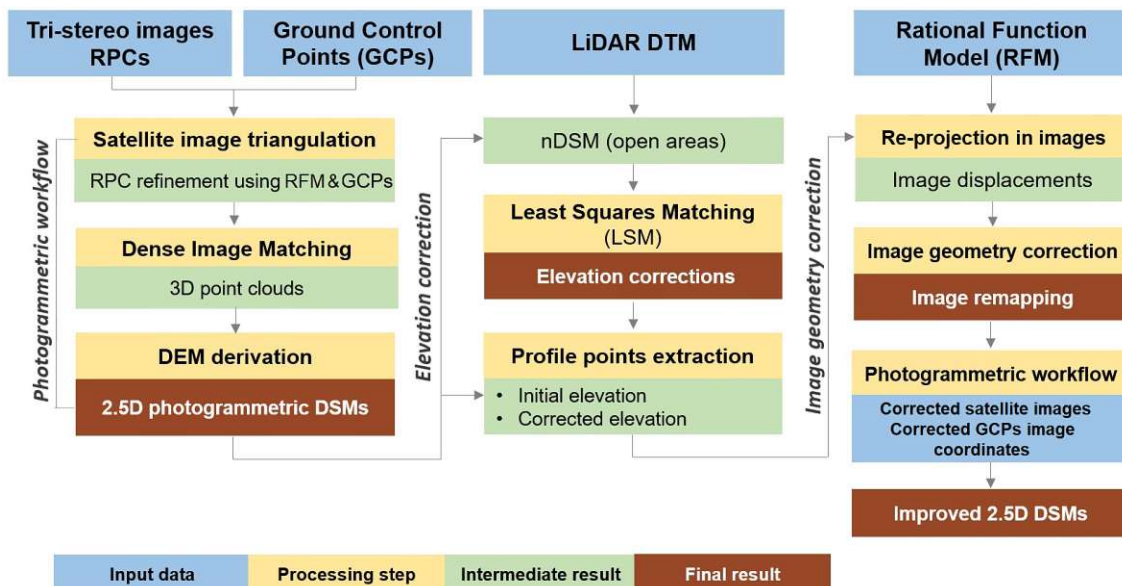3. Image geometry improvement

Figure 2.3: Processing workflow for satellite image geometry correction model.

The first part describing the photogrammetric processing chain with bias-compensation of the RPCs using GCPs for deriving elevation models from satellite imagery is detailed in section 2.1.

Secondly, the elevation corrections in object space are determined by applying the continuous LSM photogrammetric technique between satellite-based DSMs and the reference Lidar DTM in open, smooth areas. The continuous mode option of LSM uses small patches arranged along flight direction with the patch length as the dimension measured along flight direction and the patch width represented by the dimension measured perpendicularly to the flight direction. For modelling the periodic systematic elevation errors in object space, the continuous mode option of LSM was used with a patch length of 1200 pixels and 70% overlap between succesive patches. Finally, an individual transformation parameter set (offsets in Z direction) is estimated for each patch. Hence, the elevation corrections describe the offsets in Z direction as a function of Y coordinate (in flight direction):

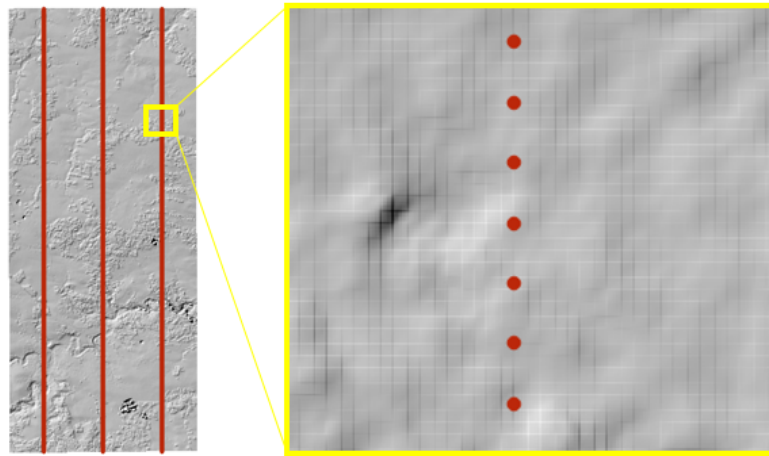$$\delta_Z = f_{LSM}(Y) \tag{2.6}$$

Figure 2.4: Definition of 3D point profile-lines at 10 m interval.

Thirdly, the image geometry corrections are determined by adopting a *"from object to image space"* re-projection procedure. The image correction starts with the extraction of 3D point profile lines. The selected points are distributed in three parallel lines at 10 m interval along flight direction (Figure 2.4). Their corresponding elevations (Z) are extracted from the photogrammetric satellite-based pairwise DSMs (FB, FN, NB) and the corrected elevations are computed by using the function of Z differences resulted from LSM-continuous approach with the following formula:

$$Z_{corr} = Z - f_{LSM}(Y) \tag{2.7}$$

where $f_{LSM}(Y)$ is the correction function for Z elevation differences in Y-coordinate direction (as resulted from continuous LSM for each DSM model: FB, FN, NB).

Starting with the normal photogrammetric processing with bias-compensation of the RPCs using GCPs and followed by the elevation corrections computation, the proposed image geometry correction method is based on the main principle *"from object to image space"*. This consists in the re-projection of the 3D points from line profiles into image space by using the RPC model (Figure 2.5).

For each point position in 2D different elevations are available: elevations extracted from the pairwise satellite-based DSMs (FB, FN, NB) and the corrected elevation
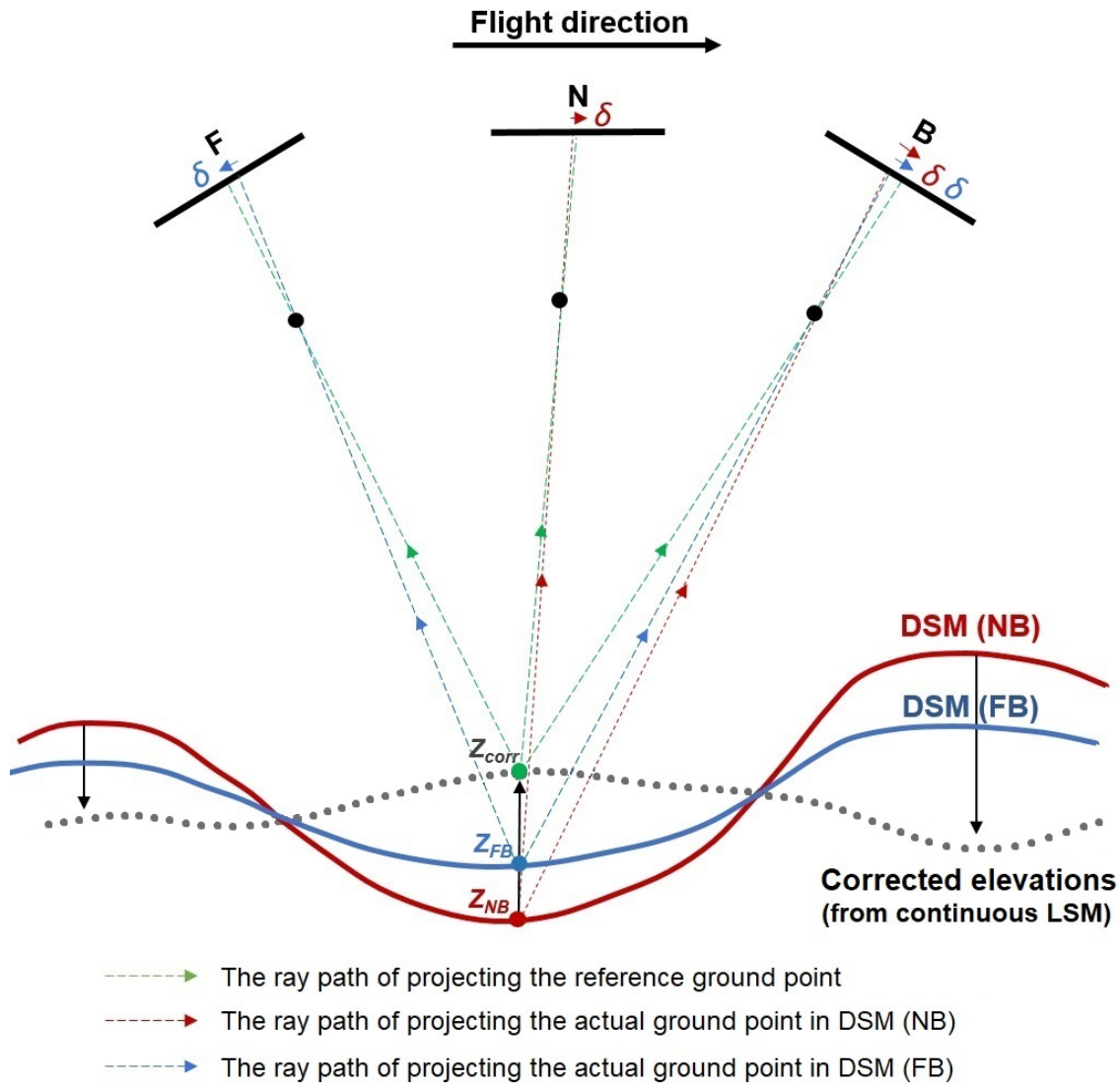
Figure 2.5: Computation of image corrections.

(computed in a previous step by applying the elevation correction function resulted from continuous LSM). By re-projection in image space, the points have different positions in the satellite images and thus, the resulting displacements serve as basis for the computation of image corrections, by using the following general formula:

$$
\begin{aligned}
(X, Y, Z) &\rightarrow (c, r) \\
(X, Y, Z_{corr}) &\rightarrow (c_{corr}, r_{corr}) \\
\delta &= r_{corr} - r
\end{aligned}
\tag{2.8}
$$

where (X, Y, Z) - 3D point coordinates with Z extracted from satellite-based elevation model showing systematic errors, (r, c) - row, column image coordinates, (X, Y, Z_{corr}) - 3D point coordinates with corrected Z elevation, (c_{corr}, r_{corr}) - corrected point position in image space, δ - image correction (displacements, computed as difference between the corrected and initial row point position in image space).

In contrast to stereo, the tri-stereo satellite image acquisition allows the computation of pairwise DSMs and therefore, each satellite image is involved in the derivation of

two pairwise DSMs: Forward image in DSM_{FB}, DSM_{FN}, Nadir image in DSM_{FN}, DSM_{NB}, and Backward image in DSM_{NB}, DSM_{FB}. Hence, two different correction functions per satellite image are computed, which are averaged in a final stage. Additionally, the distribution of points in three profile lines requires the estimation of average corrections. For example, the computation of satellite correction function for the Backward image, implies the following calculations:

(1) Re-projection of 3D points in image space

$$(X, Y, Z_{FB}) \rightarrow (c_{FB}, r_{FB})$$
$$(X, Y, Z_{NB}) \rightarrow (c_{NB}, r_{NB})$$
$$(X, Y, Z_{corr}) \rightarrow (c_{corr}, r_{corr})$$

(2.9)

where (X, Y, Z_{FB}) and (X, Y, Z_{NB}) - 3D coordinates of points extracted from satellite-based elevation models, (c_{FB}, r_{FB}) and (c_{NB}, r_{NB}) - image coordinates in Backward scene, (X, Y, Z_{corr}) - 3D point coordinates with corrected Z elevation, (c_{corr}, r_{corr}) - corrected point position in Backward image.

(2) Corrections in image space

$$\delta_{FB_i} = r_{corr} - r_{FB}$$
$$\delta_{NB_i} = r_{corr} - r_{NB}$$

(2.10)

where δ_{FB_i} and δ_{NB_i} - image corrections (displacements computed as difference between the corrected and initial row point position in image space) for each profile point line, with $i = 1, 2, 3$ - left, middle and right profile.

(3) Correction average of profile point lines

$$\delta_{FB} = \frac{\sum_{i=1}^{3} \delta_{FB_i}}{3}$$
$$\delta_{NB} = \frac{\sum_{i=1}^{3} \delta_{NB_i}}{3}$$

(2.11)

(4) Final average correction model for Backward satellite image

$$\delta_B = \frac{\delta_{FB} + \delta_{NB}}{2}$$

(2.12)

The implementation of the computed correction models to each satellite image implies a replacement of pixels from one location in the source-image to the correct location in the final image, process called warping. This strategy is possible by using the specific **remap()** function from the OpenCV library. Since a perfect one-to-one-pixel correspondence between source and destination images is not achievable, a cubic interpolation was set for non-integer pixel locations. The correction for every pixel location (c, r) is expressed as:

$$g(c, r) = f(h(c, r))$$

(2.13)

where $g()$ is the remapped image, $f()$ the source image and $h(c, r)$ is the mapping function that operates on (c, r), given by:

$$\begin{aligned} h(c) &= c \\ h(r) &= r - I(r) \end{aligned} \tag{2.14}$$

where the mapping function remains constant in X-East direction, while the $I(r)$ defines the image correction function model for displacements in Y-North direction. This function is sampled over the final average correction model (Equation 9) with a pixel spacing of approx. 10 m/0.7 m.

The image correction model was not applied to the satellite images only, but also to the coordinates of the GCPs in image space:

$$(c, r) \rightarrow (c, r - I(r))) \tag{2.15}$$

2.4 Supervised machine learning for point cloud classification

In this research, the focus falls on the supervised classification method based on initial human classification and interaction. The point clouds generated through dense image matching from the tri-stereo scenes were classified using the approach described in Waldhauser et al. (2014), where they apply Classification and Regression Trees (CART) modelling for classifying point clouds from airborne laser scanning (Breiman et al. 1984), solution implemented in the software package R (Therneau et al. 1997). Even if initially developed for airborne laser scanning products, the technique was successfully applied for image matching point clouds from high-resolution aerial images, as well (Tran et al. 2018).

2.4.1 Point feature extraction

A point cloud $\mathbf{P}$ is defined as a collection of points $p_i = (x, y, z) \in P$ with three dimensional coordinates in space. The basic concept of the adopted supervised classification approach is an individual point classification, which means that each point is separately classified based on its 3D coordinates and additional descriptors. Local point distribution measures like normal vector, surface roughness, vertical point distribution, point density need to be computed and assigned as attributes to each 3D point. Therefore the derivation of appropriate features within an adequate neighbourhood area is a crucial pre-processing step for achieving good classification results.

For computing the normal vector of a point, a local surface plane is estimated by least squares interpolation of its neighbouring points selected by a K-nearest neighbors algorithm. In a second stage, the determination of a tangent plane is an eigenvalue problem, resulting in the plane parameters and the corresponding residuals, finally leading to the normal vector in that point. The quality of the plane fitting given by the root mean square orthogonal distances between the optimal estimated plane and the neighbouring points is an indicator of surface roughness.

The normal vector estimation provides three eigenvectors with their three corresponding eigenvalues ordered by size $\lambda_1 > \lambda_2 > \lambda_3$ and which are saved in the resulting structure tensor T. The geometric information encoded in T is essential

for describing the spatial point distribution in the neighborhood, needed for the class separability. The three important features derived from T are linearity (L_T), planarity (P_T) and omnivariance (O_T). The linearity feature is used to investigate how well the distribution of the points could be modelled by a 3D line. This is essential for identifying linear structures such as powerlines or any other objects with similar geometric properties. The planarity feature is strongly related to the quality of the plane fitting for normal vector estimation and describes the smoothness of the surface. Omnivariance describes the inhomogeneuous nature of point distribution across a 3D volume. This feature is mostly met for points describing trees or vegetation.

$$L_T = \frac{\lambda_1 - \lambda_2}{\lambda_1} \tag{2.16}$$

$$P_T = \frac{\lambda_2 - \lambda_3}{\lambda_1} \tag{2.17}$$

$$O_T = \sqrt[3]{\lambda_1 \lambda_2 \lambda_3} \tag{2.18}$$

Features like anisotropy, eigenentropy and curvature can also be derived from the eigenvalues of the structure tensor T.

Additional information regarding the point distribution and surface characteristics is given by features like *echoratio*, Z_{Range}, Z_{Rank}, $Z_{Normalized}$ and *PointDistance*. *Echoratio* is a measure that describes the vertical penetrability of a surface (Höfle et al. 2012), Z_{Range} marks the maximum height difference between the points in the neighbourhood, while Z_{Rank} shows the rank of the neighbouring point corresponding to its height. $Z_{Normalized}$ is the rank of the point (between 0 and 1) multiplied by the height range in the neighboorhood and the *PointDistance* is the average distance between the points in the neigboorhood.

Thus, besides their 3D coordinates (X,Y,Z), the point cloud can be augmented by additional parameters like: roughness ($NormalSigma$), normal vector ($NormalX$, $NormalY$, $NormalZ$), echo ratio (ER), Z_{Range}, Z_{Rank}, $Z_{Normalized}$, $PointDensity$, $PointDistance$, *linearity*, *planarity*, and *omnivariance*.

2.4.2 Supervised classification using machine learning

Classification of points is a significant application for the automated processing of point clouds. During this process, each point receives an individual class label, based on its characteristics and derived features.

The main idea behind the supervised classification methods is to automatically classify a large, unseen data set by using models that have been trained on a smaller, representative set of training data. Hence, for building the classification rules during the learning process, the adopted decision tree classification requires already labelled training data. This is usually achieved by manually class assignment operation. For evaluating the model performance, true labelling information for the test data set is required as well.

Both training and test data should contain carefully sampled data that spans the various classes of interest. While the training set is represented by the sample of data being employed to fit the parameters of the classifier, the test set contains the sample of data that provide an unbiased evaluation of the model fit. The test dataset is used only once the model is completely trained, for assessing the quality

performance of the classifier. The separation of data into training and testing data sets plays a significant role in the evaluation process of the classification models. The dataset split ratio depends on two aspects: firstly, on the total number of samples in the data and secondly, on the actual model that is being trained. While some models need substantial data for the learning process requiring larger training sets, simpler models with very few hyperparameters will be easy to validate and tune, so they will need reduced sizes of the validation set. In case the model has many hyperparameters, then a larger training and validation set should be used. The dataset split ratio between training and test sets is therefore quite specific to every use case.

The supervised machine learning classification model comprises the following three main steps (Figure 2.6):

(1) point feature extraction

(2) model fitting to the training data, known as the learning process

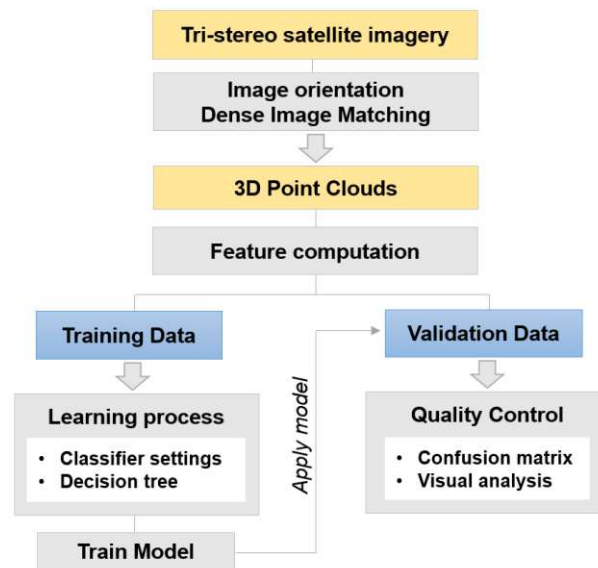(3) application of the classification model to predict the labels for the test data.

Figure 2.6: Overview of the general classfication workflow.

In the learning phase, the following information is needed: (i) list of classes, (ii) hand-crafted 3D geometric and colour features and (iii) reference labelled dataset as training data.

The chosen classification tree method is actually the predictive model that creates the links between the point features and its corresponding class. Structurally, a classification tree has two parts: leaves - which represent the final class labels for the specific point after attribute analysis, and the branches - which are the conjunctions of features that lead to the choice of the class label. The classification tree seeks to partition the entire feature space of a data set, one variable at a time, by selecting a variable and an appropriate splitting value (Waldhauser et al. 2014).

The performance of the machine learning classifier is then evaluated by building the confusion matrix and computing the Accuracy, Precision, Recall, and F1-Score metrics. Accuracy is the performance measure given by the ratio between the correctly predicted observations and total observations. Precision is the proportion of

positive identifications that are actually correct and is given by the ratio between the correctly predicted positive observations and the total predicted positive observations. Recall (sensitivity) is the proportion of actual positives that were correctly identified and is defined by the ratio of correctly predicted positive observations to the all observations in the actual class. F1-Score is the weighted average of Precision and Recall. This score takes into account both false positives and false negatives, therefore is more useful than the Accuracy metric, especially in the case of an uneven class distribution.

$$Accuracy = \frac{TruePositive + TrueNegative}{Total} \tag{2.19}$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{2.20}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{2.21}$$

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{2.22}$$

3

Materials

3.1 Study sites

Four distinct areas of interest with particular characteristics and located in different countries and continents (Austria, Spain, Sweden in Europe and Nepal in Asia) were chosen as study sites for the conducted analyses. They were selected based on two main criteria: (1) the availability of reference data, and (2) the diversity of the land cover and relief. They describe distinct types of landscapes (rural, urban, grasslands, forests) and topographic conditions (flat, hilly, and mountainous areas), which come with their own unique and specific challenges in the photogrammetric processes of dense image matching and 3D reconstruction.

The surface land coverage was visualised on the Corine Land Cover Map version 2018 from Copernicus Land Monitoring Service with an open access for all users (*https://land.copernicus.eu/pan-european/corine-land-cover*). The map is based on Sentinel-2 (GSD of 10 m) and Landsat-8 (GSD of 30 m) satellite imagery.

3.1.1 Allentsteig study site

The Allentsteig study area is located in Lower Austria (48° 30' 30"N; 15° 08' 34"E; WGS84, 33N), the north-eastern state of the country, a hilly region with elevations ranging from 537 to 846 m above sea level (Figure 3.1 first row). From 639 km^2 total extend, about 47% is covered by arable lands, grasslands and agricultural fields, while about 39% is covered by coniferous and broad-leaved forests. Urban, suburban and rural areas are also present (for e.g., Allentsteig, Zwettl cities, Niedernondorf, Scheideldorf villages), but they cover only small parts, compared with the entire area extend.

3.1.2 Barcelona and Stockholm cities

The urban landscapes selected as study cites are two capitals in Europe: Stockholm (59° 19' 46"N; 18° 04' 07"E) - capital of Sweden and Barcelona (41° 23' 02"N; 02° 11' 14"E) - capital of the autonomous community of Catalonia.

Located on northeast coast of the Iberian Peninsula, Barcelona (Figure 3.1 second column) is one of the largest metropolises on the Mediteranean Sea and the second most populous municipality of Spain (Cook 1910). As described by Marshall (2004), the location of the city on the coast brings certain advantages, like: moderate climate, ease communication for goods and many leisure opportunities. At a larger scale, it can be said that Barcelona is situated on a coastal corridor between France, Italy and Africa. At a more local scale, the position of the city has specific features:
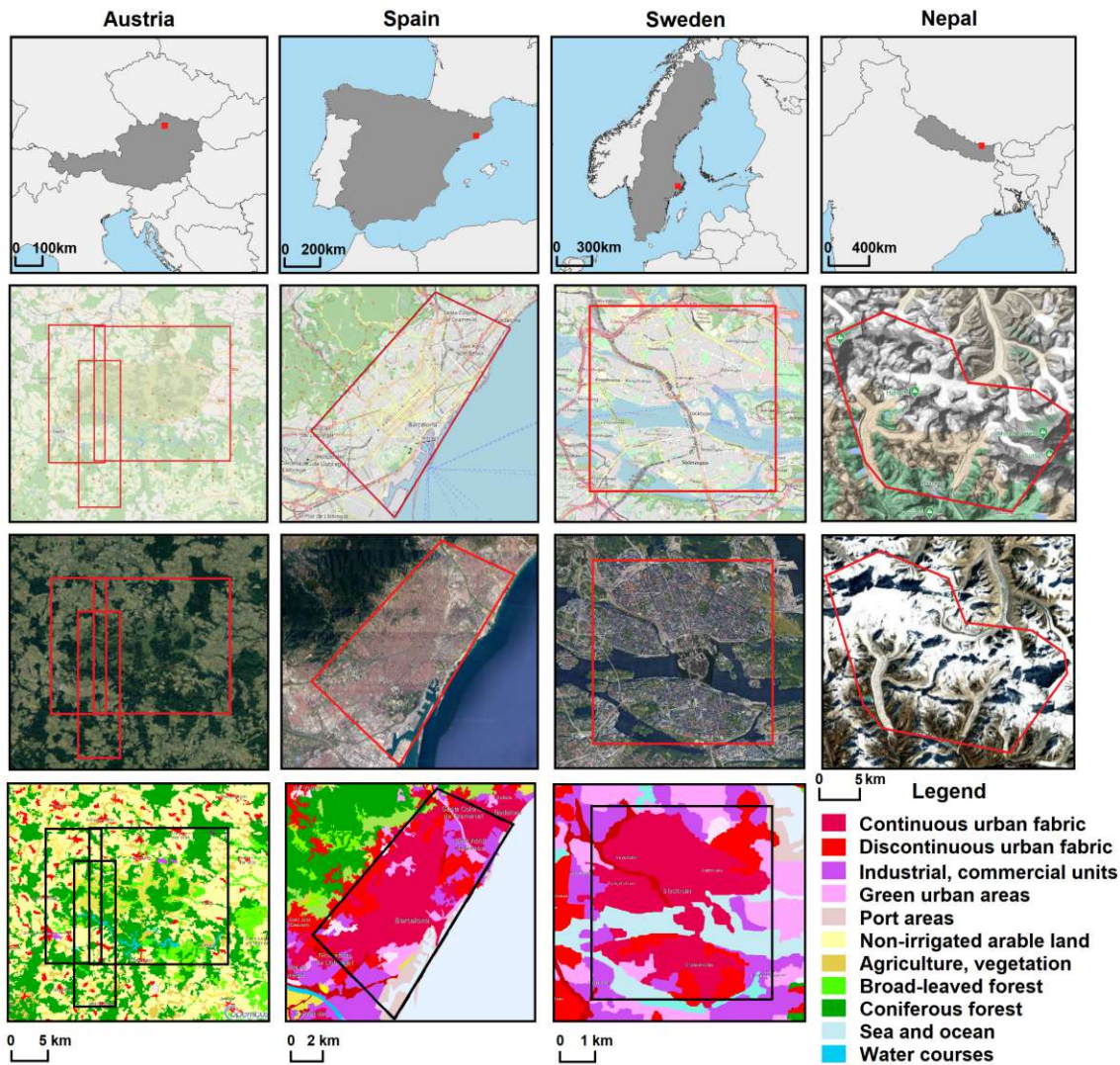
Figure 3.1: Locations of the four study areas with satellite images acquisition footprints overlaid in OSM (second row, except Himalaya topographic map), Google Earth (third row), and Corine Land Cover Map 2018 - Copernicus (fourth row).

this is actually a wide sloping site between the mouths of two important rivers, the Lobregat to the southwest and Besos to the north (Cook 1910). At the west side, Barcelona is limited by the mountain range of Collserola with the highest peak of 512 meters. Located at approximative 120 kilometers south from the Pyrenees mountains, the urban extent of the city covers an area of 101.4 km^2 (Balibrea 2001). The study area of 100 km^2 covers Barcelona city with its buildings, streetscapes, infrastructure, industrial, commercial units, green spaces, and port areas. Barcelona is a major cultural, economic, and financial centre in southwestern Europe with a population of 1.6 million within city limits (Delgado 2012).

Situated on one of the world's most beautiful harbors, Stockholm (Figure 3.1 third column) is the biggest city in Sweden and most populous urban area in Scandinavia (Griffiths 2009). Located on the east coast of the state, where Lake Mälaren - Sweden's third-largest lake- flows into the Baltic Sea, the central parts of the city stretches across fourteen islands that are continuous with the Stockholm archipelago (De Geer 1923). As described by Janson et al. (1965), with its geographical center

positioned on water, in Riddarfjärden bay, Stockholm covers a total area of 188 km^2. Over 30% of the city area consists of waterways and approximately 30% is covered by green spaces and parks. Stockholm is the natural capital of Scandinavia, due to the following reasons: in a first place, it is located in the heart of the region, benefiting from a very good transport infrastructure and in a second place, it is the largest city in the largest country in Scandinavia. Moreover, here are located the most multinational companies and the largest stock market (Janson et al. 1965). Stockholm Municipality is the cultural, political, media, and economic centre of Sweden, with a population of 965.232 inhabitants (as statistics results of 2017 show).

3.1.3 Himalaya mountains

Himalaya, the world's highest and most spectacular mountain range in Asia separates the plains of the Indian subcontinent from the Tibet region in China (Figure 3.2a). The name "Himalayas" comes from the Sanskrit term with "Hima" meaning snow and "Layas" meaning residence or hometown (Yang et al. 2004). Thus, the Himalayas are the hometown of snow. With elevation averages of 6 200 meters, the mountain range runs northwest to southeast in an arc of 2 400 kilometers long and 200 to 300 kilometers wide. Here are more than 40 mountains over 7 000 meters high and nine of the world's ten highest peaks, including Mount Everest (8 848 m a.s.l.) at the border between Nepal and China (Pratt 1855). Home to around 15 000 glaciers, the Himalaya contain the third largest deposit of ice and snow on the planet, after Antarctic and Arctic. Moreover, many of Asia's most important rivers have their sources in the Himalaya Mountains including the Indus, Ganges, and Brahmaputra (Yang et al. 2004).

The study area with an extend of 500 km^2 is located in Khumbu (region of Nepal, at 28° 03' 48" N; 86° 42' 24" E), a section of the Himalayas in northeast Nepal and south-central Tibet of China, which includes the following most important peaks: Everest (the Earth's highest mountain, 8 849 m), Lhotse (the fourth highest mountain in the world with 8 516 m), Nuptse (7855 m), Lobuche (6119 m), Cho Oyu (8 188 m), Gyachung Kang (7 952 m), and Pumori (7 161 m) (Burrard 1907). More than 50% of the area is covered by the following glaciers: Ngozumpa (in the northwest side), Pumori (in the central part), and Nuptse and Lhotse (in east). An overview with the ice-covered peaks and glaciers in the acquired Pleiades imagery from october 2015 is given in Figure 3.2b.
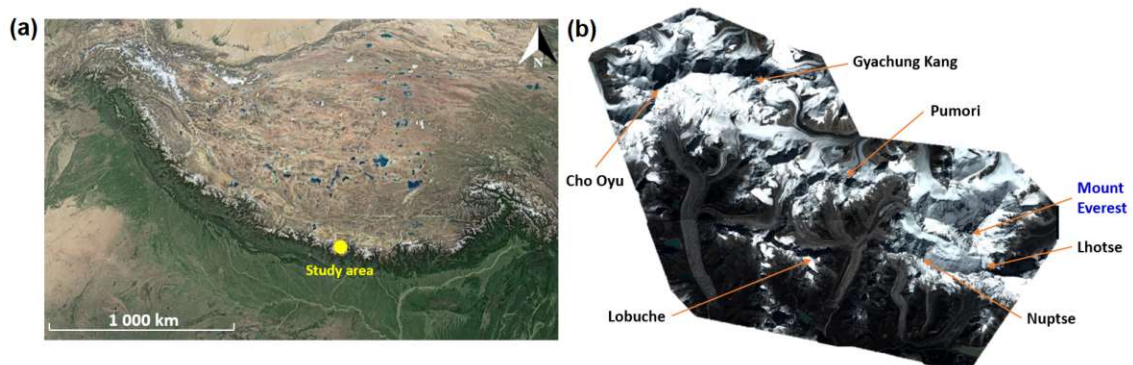

Figure 3.2: (a) Himalaya range mountains with location of the study area; (b) overview of Pléiades satellite imagery from october 2015.

3.2 VHR Satellite Imagery Datasets

3.2.1 Pléiades and WorldView-3 VHR satellite sensors

Analyses were conducted based on stereo and tri-stereo satellite images acquired with two VHR optical satellite sensors, Pléiades and WorldView-3 .

The Pléiades satellite constellation comprises two identical satellites Pléiades 1A and Pléiades 1B, which were launched by the French space agency (CNES - Centre National d'Etudie Spatiales) in December 2011 and December 2012, respectively. Both satellites are flying on the same sun-synchronous low-Earth orbit at an altitude of 694 km with an orbital period of 98.8 minutes and an offset of 180° from each other, which provides a daily revisit interval (Zhou et al. 2015). For the tri-stereo acquisitions, the B/H ratios (baseline/height) vary between 0.1 and 0.5 with corresponding stereo angles of 6° to 28° (De Lussy et al. 2012). The sensors are able to collect both panchromatic imagery with 50 cm resolution and four-band multispectral imagery with 2 m resolution (Astrium 2012).

WorldView-3 Digital Globe's very high-resolution optical sensor was launched in August 2014, having an estimated service life time of 10 to 12 years. The satellite operates at an altitude of 617 km and has an orbital period of 97 minutes with an average revisit time of less than 1 day. With a telescope primary mirror featuring an aperture diameter of 110 cm, the sensor is able to collect panchromatic imagery with 31 cm resolution, eight-band multispectral imagery with 1.24 m resolution, eight-band shortwave infrared imagery at 3.7 m resolution and 12 CAVIS-bands (Clouds, Aerosols, Vapors, Ice, and Snow) data at 30 m resolution (DigitalGlobe 2012). Similar to the Pléiades system, the WorldView-3 sensor uses the pushbroom scanning technique with a 35 000 pixel dectector array for panchromatic and a 9 300 pixel detector array for multispectral bands, both arranged perpendicular to the flying direction and collecting images one line at a time (pixels in line are measured simultaneously).

3.2.2 Satellite image acquisition for Allentsteig study site

For the Allentsteig study area the analyses were conducted based on tri-stereo satellite images acquired with both Pléiades-1B and WorldView-3 optical satellite sensors. Each triplet consists of images that were acquired during a single pass, from different along-track positions of the satellite: forward (F), close to nadir (N) and backward view (B).

The first Pléiades tri-stereo dataset were acquired in the late morning of June 13, 2017, around 10:10 a.m in North-South direction. Due to the high speed ($\approx$ 7.5 km/s) the sensor mounted on PHR-1B platform was able to cover 159 km^2 within 23 seconds. The along-track viewing angles w.r.t. nadir are -5.66^o, 0.46^o and 5.19^o for the forward, close to nadir and backward images, respectively.

A week later, on June 20, 2017 the same platform acquired 383 km^2 (within 20 seconds), with an overlap of 45 km^2 with the first dataset. The corresponding along-track viewing angles w.r.t. nadir are of -5.49^o, 0.06^o and 6.09^o.

To comparatively investigate the accuracy potential of DEM derived from VHR satellite imagery, the impact of the acquisition geometry with a different GSD, and the classification performance, we tasked a new tri-stereo WorldView-3 dataset in

the same area (100 km^2), with an overlapping of 44.5 km^2 with the Pléiades images. According to the metadata, the tri-stereo WorldView-3 images were acquired on April 8, 2018, around 10:22, within 37 seconds. The along-track acquisition angles w.r.t. nadir are of -11°, -0.6^o and -12.2^o for the forward, close to nadir and forward images, respectively.

The spatial extend of the Pléiades tri-stereo images is of 19 x 8.35 km (158.7 km^2) for the first acquisition and of 19 x 20.15 km (383 km^2) for the second data set, respectively. The WorldView-3 data coverage has a total extension of 19 x 5.3 km (100 km^2). The location of the study area with the three acquisitions and marked overlapping surface of 44.5 km^2, is shown in Figure 3.3.
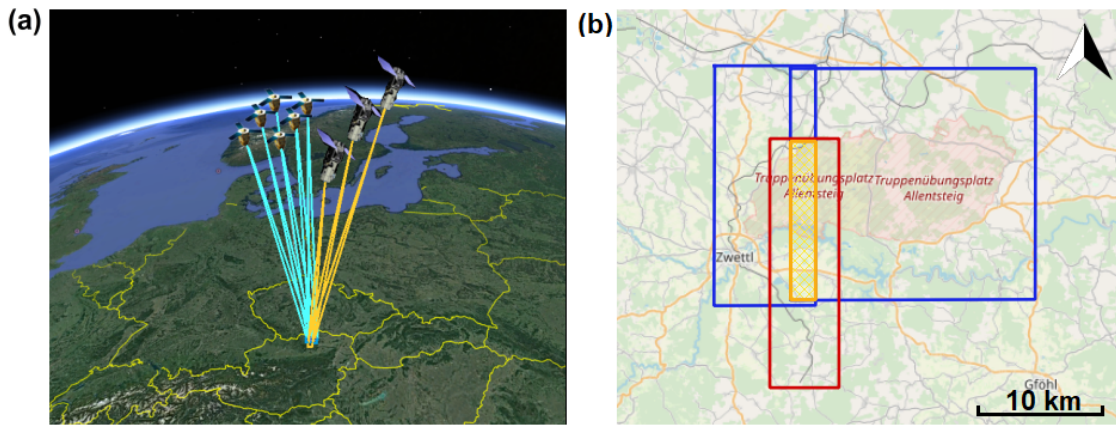

Figure 3.3: Allentsteig study area: (a) Acquisition geometries of Pléiades (blue) and WorldView-3 sensors (orange) (view of the satellite positions in Google Earth); (b) acquisition footprints with marked common area (yellow) overlaid on OSM.

Detailed information regarding the acquisition properties for all the tri-stereo datasets are shown in Table 3.1 (first three rows). The stereo intersection angles (also called convergence angles in Karsten Jacobsen and Topan 2015) between each two scenes and accordingly their base-to-height (B/H) ratios were computed using the equations found in (Perko, Raggam, Schardt, et al. 2018). Table 3.1 summarizes the main values for the viewing angles along- and across-track direction w.r.t. the nadir, GSD, B/H ratio and convergence angles on the ground. For the Forward-Backward (FB) image combination the baseline to height ratios (B/H) are of 0.24, 0.25 and 0.45 for the three acquisitions, with the corresponding convergence angles of 13.8^o, 13.9^o, and 25.5^o. The acquisition geometries of both Pléiades tri-stereo datasets are very similar w.r.t the values for the B/H ratio and convergence angles on the ground. In contrast, the WorldView-3 acquisition shows values approximately two times larger. This is because the WorldView-3 platform flies at a lower altitude (617 km), fact that leads to higher values for the B/H ratios, even if the satellite speed is almost the same. Thus, the B/H ratios for the FN and NB are both 0.22, whereas for FB is 0.45. The intersection angles of rays on the ground for each combination have higher values as well: 12.7^o (FN), 12.8^o (NB) and 25.5^o (FB).

All optical satellite images were delivered as 16 bit pansharpened with four spectral bands (Red, Green, Blue and Near-infrared). With radiometric resolutions of 16 bit/pixel, the images provide a higher dynamic range than the traditional 8- or 12-bit/pixel images. Depending on the viewing angle, the spatial resolutions

(mean values for the GSD) for Pléiades images are varying between 0.70 and 0.71 m, whereas the WorldView-3 sensor provides images with higher resolutions, between 0.31 and 0.33 m. For each satellite image, the RPCs, allowing the conversion between image and object space (Fraser, Dial, et al. 2006), were provided by the supplier.

The tri-stereo Pléiades images were provided at sensor processing level, corrected only from on-board distortions such as viewing directions and high frequency attitude variations (Panem et al. 2012). In contrast, the WorldView-3 images were delivered as tri-stereo with relative radiometrically-corrected image pixels. The relative radiometric calibration includes a dark offset subtraction and a non-uniformity correction (e.g. detector-to-detector relative gain) which is performed on the raw data during the early stages of product generation. The WorldView-3 images were delivered as eight, two and two tiles for nadir, forward and backward, respectively. As a pre-processing step, the tiles for each image were mosaicked accordingly.

An important characteristic of both satellites – Pléiades-1B and WorldView-3 – is the fast rotation to enable recording three images within the same orbit in less than 1 minute. Hence, the operational aspects of acquisition influence the radiometric characteristics of the images. The fast recording (less than one minute) guarantees same sun illumination conditions with no significant shadow changes in the scenes. From a visual inspection, some radiometric inconsistencies can be observed in the WorldView-3 forward image, i.e. saturation and spilling effects (Figure 3.4), mainly caused by the reflective roof surface, in combination with the imaging incidence angle. In contrast, no radiometric artefacts were observed in the Pléiades satellite images.
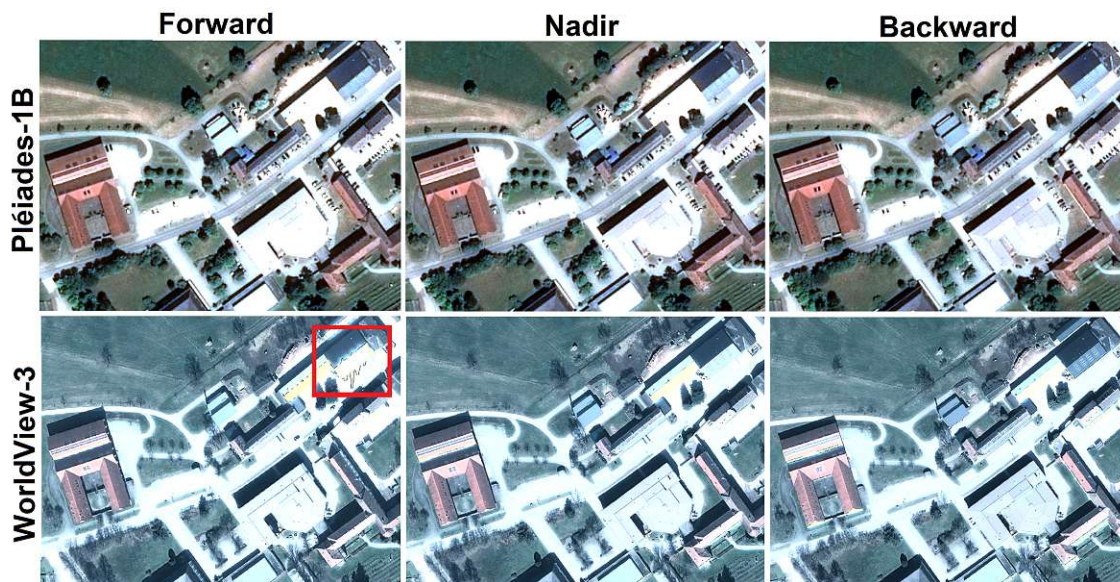

Figure 3.4: Detail of Pléiades (first row) and WorldView-3 (second row) tri-stereo satellite images on the same built-up area, acquired with forward-, nadir- and backward-looking.

Figure 3.5 shows a visual comparison between the pan-sharpened (near) nadir images acquired with Pléiades (left) and WorldView-3 (right) over a small area, in order to highlight the effects of the two different resolutions of 0.70 m and 0.31 m. The same

objects can be distinguished in both images: streets, buildings, cars and trees. The higher resolution of WorldView-3 provides a more detailed image with clear visible cars, tree branches and building roof edges.
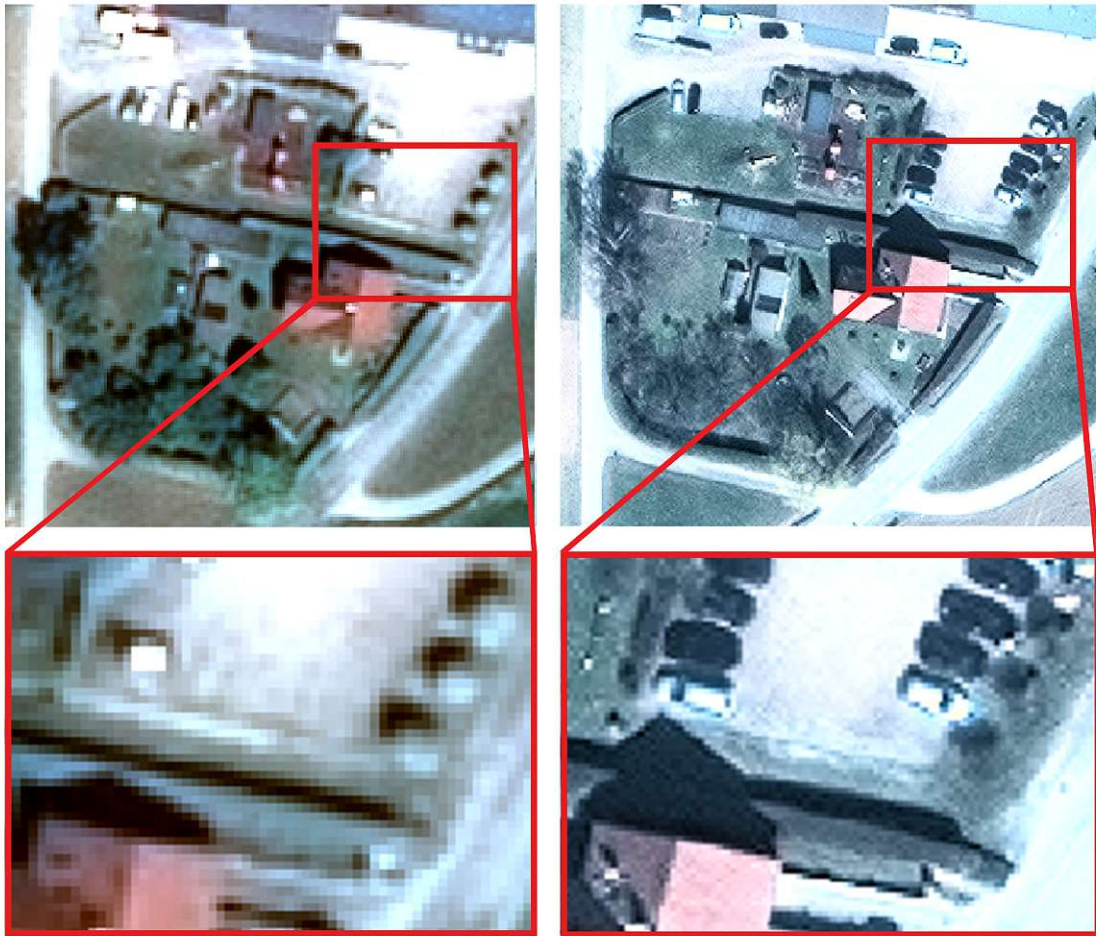

Figure 3.5: Comparative view of the same area from Pléiades (left) and World View-3 images (right) with a magnified detail.

3.2.3 Satellite imagery for Barcelona and Stockholm cities

Both satellite imagery datasets corresponding to the two municipalities of Barcelona and Stockholm are acquired by the VHR WorldView-3 sensor in stereo mode, in less than one minute.

The pansharpened stereo pair for Barcelona was acquired in summer, on July 20, 2017 and covers $100\ km^2$. The first image was collected in forward mode with an off-nadir viewing angle along track of 9.41^o and the second image in backward looking mode with a corresponding angle of -19.70^o. The mean spatial resolutions (GSD) are of 0.34 and 0.35 cm, the B/H ratio is equals to 0.71 and the convergence angle on the ground has a value of 38.7^o. For both images, the mean sun elevation is of 66.5^o.

The pansharpened stereo satellite images for Stockholm were acquired during the autumn season, in the late morning of November 17, 2017. The collected surface of $27\ km^2$ covers the main down-town area of the city center. With 0 cloud coverage,

the images have mean spatial resolutions of 0.35 and 0.40 cm, a B/H ratio of 0.73, and a convergence angle on the ground of 36.3^o. Images are taken in the leaf-off season and suffer from poor illumination due to the low elevation angle of the sun (11.55^o for the forward and 11.51^o for the backward image). To improve the visual impression, the grey (color) values can be stretched. Independent thereof, dark shadow areas and large shadow areas prevail and render the analysis of image content difficult.

The optical satellite images for the two cities were delivered as OR2A (Ortho-Ready Standard 2A) products with included Rational Functional Model (RFM), used for replacing the physical sensor model. The OR2A Imagery are map projected products (UTM WGS84), but no topographic relief is applied to the reference ellipsoid. Instead, the imagery is only projected to a constant base elevation. Thus, this product is suitable for orthorectification. All images are geometrically and radiometrically corrected. An overview of the acquired satellite imagery for the two cities are shown in Figure 3.6.
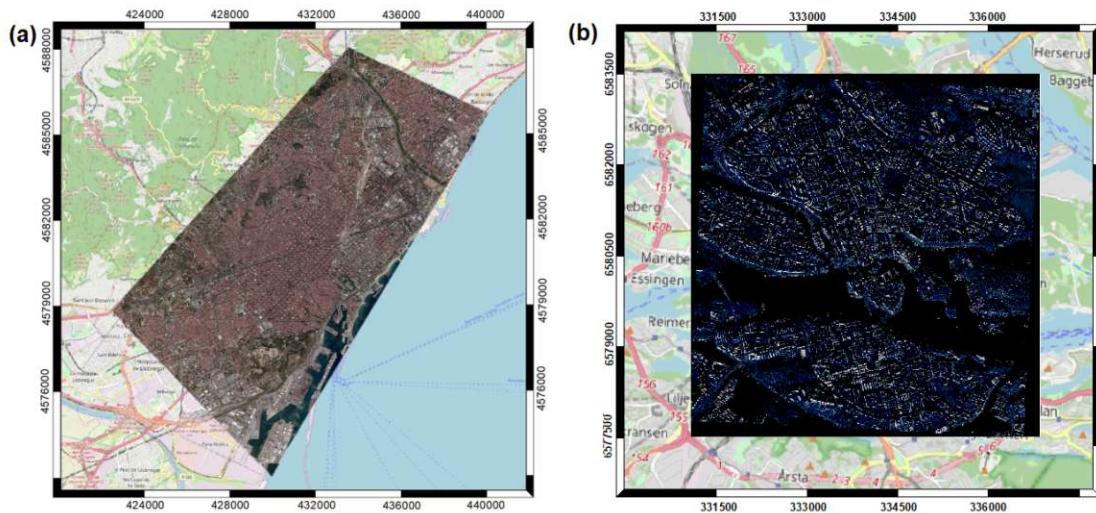

Figure 3.6: Overview of the WorldView-3 satellite imagery for the urban areas of interest: (a) Barcelona (coordinates in UTM zone 31N); (b) Stockholm (coordinates in UTM zone 34N) (It is to be noted that the WorldView-3 satellite imagery are delivered as map-oriented products, at Ortho Ready 2A processing level).

All the images are pan-sharpened with four bands (blue, green, red, and near-infrared), with spatial resolutions that variy between 0.34 m and 0.40 m, depending on the viewing angle. The stereo products allow the extraction of 3D information and derivation of DEMs through image matching techniques and photogrammetric processing. Detailed information about the two datasets are shown in Table 3.1 (rows 4 and 5).

3.2.4 Satellite imagery for Himalaya mountains

Three Pléiades triplets were provided by CNES (Centre National d'Etudes Spatiales) for the Himalaya study area: first acquisition in spring 2015, second in autumn 2015, and the third one in spring 2016. Every triple consisting of a forward, nadir, and backward scan, was taken within less than one minute from the same orbit. An

overview of the three acquisitions is given in Figure 3.7 and the main acquisition parameters are summarized in Table 3.1 (last three rows).

Figure 3.7: Overview of the Pleiades satellite imagery for Himalaya mountains at different acquisition times: (a) spring 2015; (b) autumn 2015; (c) spring 2016.

The satellite images were delivered as panchromatic, at sensor processing level, corrected from on-board distortions like viewing directions and high frequency attitude variations. The values for the GSDs vary between 0.69 and 0.73, with the lowest values corresponding to the close to nadir images. The total covered surface is of 494.13 km^2. The values for the B/H ratios and convergence angles on the ground are very similar between the three acquisitions, with variations of up to 10%. Depending on the acquisition times, the sun elevations have values of 62.39^o in spring 2015, 50.17^o in autumn 2015 and of 57.34^o in spring 2016.

Table 3.1: Acquisition properties of Pléiades and WorldView-3 tri-stereo and stereo satellite images for Allentsteig, Barcelona, Stockholm, and Himalaya study sites.

Study Area / Sensor Type / Acq. date	View	Acq. Time (h:m:s)	Viewing Angles (°) Along	Across	Overall	GSD (m)	B/H Ratio	Convergence Angles(°)	Sun Angles (°) Azimuth	Elevation	Area (km^2)
Allentsteig Pléiades 13-06-2017	Forward (F)	10:09:51.5	5.68	2.95	6.39	0.71	0.13 (FN)	7.5 (FN)	154.04	62.79	158.73
	Nadir (N)	10:10:00.9	0.49	3.17	3.21	0.70	0.11 (NB)	6.3 (NB)	154.52	62.86	158.49
	Backward (B)	10:10:11.0	-5.17	3.42	6.19	0.71	0.24 (FB)	13.8 (FB)	154.52	62.86	158.78
Allentsteig Pléiades 20-06-2017	Forward (F)	10:06:05.8	5.59	-1.67	5.84	0.71	0.13 (FN)	7.5 (FN)	152.06	62.69	383.22
	Nadir (N)	10:06:15.1	0.41	-1.44	1.50	0.70	0.12 (NB)	6.4 (NB)	152.06	62.69	377.97
	Backward (B)	10:06:25.3	-5.18	-1.19	5.31	0.71	0.25 (FB)	13.9 (FB)	152.06	62.69	380.22
Allentsteig WorldView-3 08-04-2018	Forward (F)	10:22:07.0	11.02	7.70	13.50	0.32	0.22 (FN)	12.7 (FN)	165.50	48.00	100.00
	Nadir (N)	10:22:25.5	-0.62	7.23	7.30	0.31	0.22 (NB)	12.8 (NB)	165.60	48.10	100.00
	Backward (B)	10:22:44.1	-12.21	6.72	13.91	0.31	0.45 (FB)	25.5 (FB)	165.80	48.11	100.00
Barcelona WorldView-3 20-07-2018	Forward (F)	11:04:28.3	14.41	11.40	18.22	0.34	-	-	147.31	66.52	100.00
	Backward (B)	11:05:24.5	-19.70	9.59	21.82	0.35	0.71 (FB)	38.7 (FB)	147.82	66.51	100.00
Stockholm WorldView-3 17-11-2017	Forward (F)	10:49:30.6	10.92	-19.32	22.11	0.35	-	-	184.00	11.71	27.00
	Backward (B)	10:50:30.5	-23.12	-20.24	30.32	0.40	0.73 (FB)	36.3 (FB)	184.31	11.72	27.00
Himalaya Pléiades 08-04-2015	Forward (F)	04:59:38.6	15.17	-3.20	15.48	0.73	0.23 (FN)	13.2 (FN)	135.65	62.39	494.13
	Nadir (N)	05:00:01.4	2.80	-2.45	3.72	0.69	0.25 (NB)	14.1 (NB)	136.07	62.54	563.47
	Backward (B)	05:00:24.6	-10.15	-1.67	10.29	0.71	0.48 (FB)	27.3 (FB)	136.07	62.54	550.45
Himalaya Pléiades 19-10-2015	Forward (F)	05:07:27.0	10.94	11.16	15.44	0.73	0.26 (FN)	14.1 (FN)	160.06	50.17	572.68
	Nadir (N)	05:07:51.3	-2.51	11.95	12.19	0.70	0.22 (NB)	12.6 (NB)	160.05	50.17	591.37
	Backward (B)	05:08:05.1	-10.21	12.36	15.83	0.73	0.48 (FB)	26.7 (FB)	160.42	50.24	572.08
Himalaya Pléiades 24-03-2016	Forward (F)	04:59:04.3	15.69	-3.58	16.05	0.73	0.25 (FN)	14.3 (FN)	140.45	57.34	540.05
	Nadir (N)	04:59:28.0	2.80	-2.80	3.96	0.69	0.25 (NB)	14.1 (NB)	140.45	57.34	608.40
	Backward (B)	04:59:51.4	-10.21	-2.02	10.40	0.71	0.50 (FB)	28.4 (FB)	140.45	57.34	602.58

3.3 Reference Data

For the Allentsteig study site in Austria different reference data are available: GCPs, DTM from LiDAR data, and a digital orthophoto. The 43 GCPs were measured by means of Real Time Kinematic (RTK) with high accuracy of ≈ 1 cm (Figure 3.8). Besides the improvement of georeferecing, the GCPs are used for the punctual accuracy analysis of the photogrammetrically derived elevation models. For the open, free area accuracy analyses, the DTM from lidar was employed. With spatial resolution of 1 m, the DTM was derived from an ALS flight measurement campaign in December 2015. Its vertical accuracy was checked against the RTK GCPs showing a σ_Z of 0.12 m. Besides accuracy assessment, the lidar terrain model was used to compute a reference mask for the open, smooth surfaces, to generate the Pléiades and WorldView-3 nDSMs and to improve the absolute geolocation of the photogrammetrically derived DEMs.
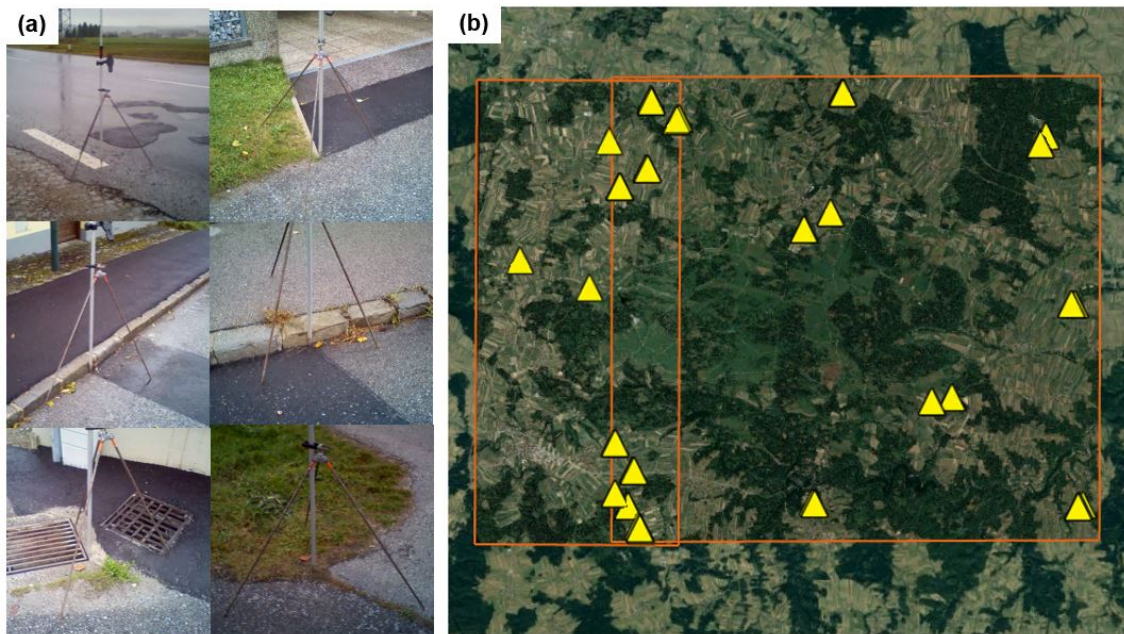

Figure 3.8: Measurement of GCPs during the RTK GNSS field survey on October 23, 2017: (a) Photos with point locations; (b) Overview of the GCPs distribution in Google Earth.

The planimetric accuracy of the available digital orthophoto from 2017 at 0.20 m spatial resolution was checked by computing the differences between the RTK point coordinates and their corresponding position in the orthophoto. The obtained result showed no shifts larger than one pixel. The digital orthophoto was used for defining the positions of new ground control and check points, whose corresponding heights were extracted from the reference lidar DTM, at the same location. The locations of the newly defined GCPs and CPs were selected in such a way that they are visible in both orthophoto and terrain model, they are identifiable in all the three satellite images, and are homogenously distributed over the entire area of interest. Moreover, the new points are located at the ground level, describing stable features in time with a high color contrast in the images. Hence the new manually measured GCPs and CPs are located at street marks (e.g. pedestrian crossing lines), intersections

and road surface changes, corners and intersections of sidewalks, corners of paved areas and parking lots, parking space lines and intersections of these lines. For the Allentsteig study area (first acquisition), besides the 31 RTK GCPs, there were manually measured 12 GCPs and 50 CPs (Figure 3.9). Additionally, to study the impact of the number and distribution of GCPs on DSM accuracy, a total number of 300 GCPs homogenously distributed over the Allentsteig scene were used (the previously defined orthophoto 12 GCPs were kept and an additional number of 288 orthophoto GCPs were measured) (Figure 2.2).
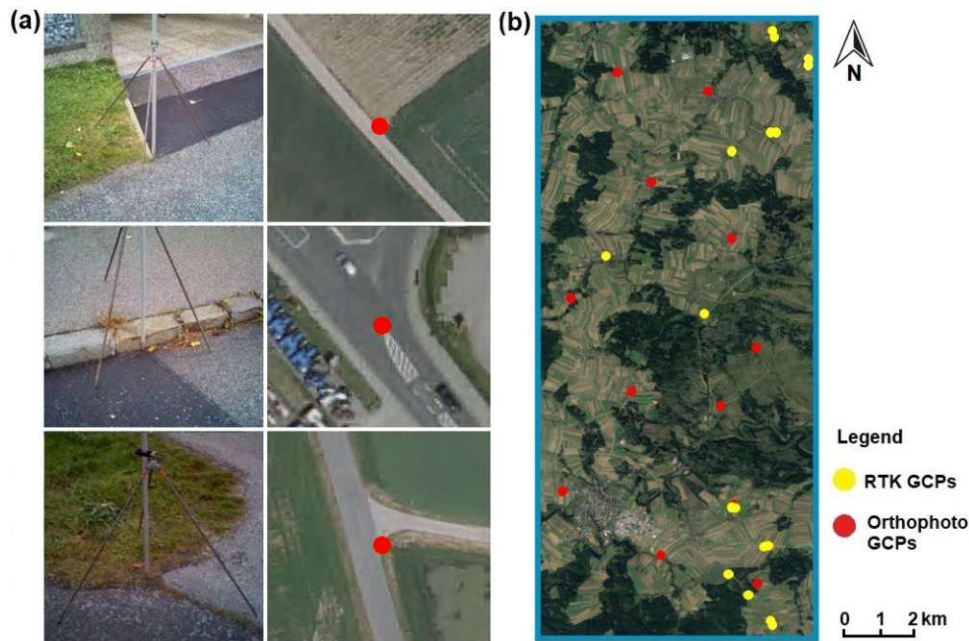

Figure 3.9: Manual measurement of GCPs and TPs: (a) Point identification in the reference othophoto, lidar DTM, and measurement in satellite image (first column) and examples of GCPs (second column); (b) Overview of the measured GCPs and CPs distribution in Google Earth.

The reference data for Barcelona city consists of an orthophoto with 0.25 m spatial resolution and a LiDAR point cloud acquisition from 2016, with 0.5 points/m^2 density. These datasets were downloaded from the publicly available database of the Cartographic and Geological Institute of Catalonia, Spain (ICGC - Institut Cartogràfic i Geològic de Catalunya) at the following web address: *https://www.icgc.cat/en/Downloads*. Due to the low point density, the lidar point clouds over Barcelona were interpolated into a 2 m resolution surface model. This was further used for elevation extraction of 23 GCPs, whoose planimetric position was previously defined in the orthophoto. The manually measured GCPs are located at ground surface level, describing stable features like street marks, corner intersections of pavement and green areas, football field marks and road surface changes (Figure 3.10). With a homogenous distribution over Barcelona, the defined GCPs were used for a planimetric geolocation assessment of the initial imagery products at OR2A processing level, as well as for the height accuracy assessment of the photogrammetrically derived surface model over the city.

For Stockholm and Himalaya the ALOS Global Digital Elevation Surface Model "ALOS World 3D - 30m" (AW3D30) was used for comparisons. This dataset
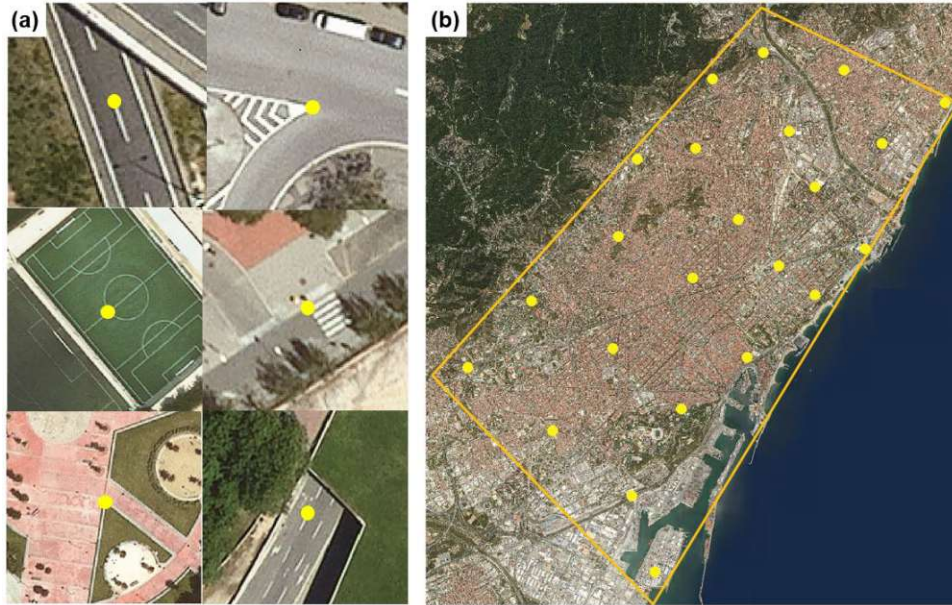
Figure 3.10: (a) Examples of measured GCPs in the othophoto; (b) Overview of the GCPs distribution.

makes part from the January 2021 version of the global digital surface model, freely available for download at *https://www.eorc.jaxa.jp/ALOS/en/dataset/aw3d30/aw3d30_e.htm*, after a pre-registration. Acquired with the Panchromatic Remote-sensing Instrument for Stereo Mapping (PRISM), an optical sensor on board of the Advanced Land Observing Satellite "ALOS", the model has approximately 30 m horizontal resolution (1 arcsecond) and a target elevation accuracy of 5 m (RMS). Even if with a coarser resolution, comparisons between the different data sources brought new insights about the capabilities of the VHR optical sensors. Besides this, comparisons between different data acquisitions over same area provide valuable information for satellite image orientation, dense matching and computation of digital elevation models.

Additionally, for the two cities - Barcelona and Stockholm - the freely available Open Street Map (OSM) geodatabase was used as a reference for the planimetric geo-location accuracy analysis. Besides the map, the datashape files from *http://download.geofabrik.de/* were used. These contain vector and metadata information for buildings, traffic, streets, railways, green spaces, and waterways.

To be mentioned that for the investigation of small isolated objects, additional manual reference measurements were done for height extraction and object size classification. Since these were employed in a single analysis of the thesis, their detailed description is found in the corresponding section 4.6.1.

<div style="text-align: right">**4**</div>

Experimental analyses and results

Present chapter contains the main results from the performed experimental analyses, starting with image orientation and accuracy of map projected satellite images (section 4.1) followed by 3D reconstruction and DSM derivation from VHR satellite imagery (section 4.2).
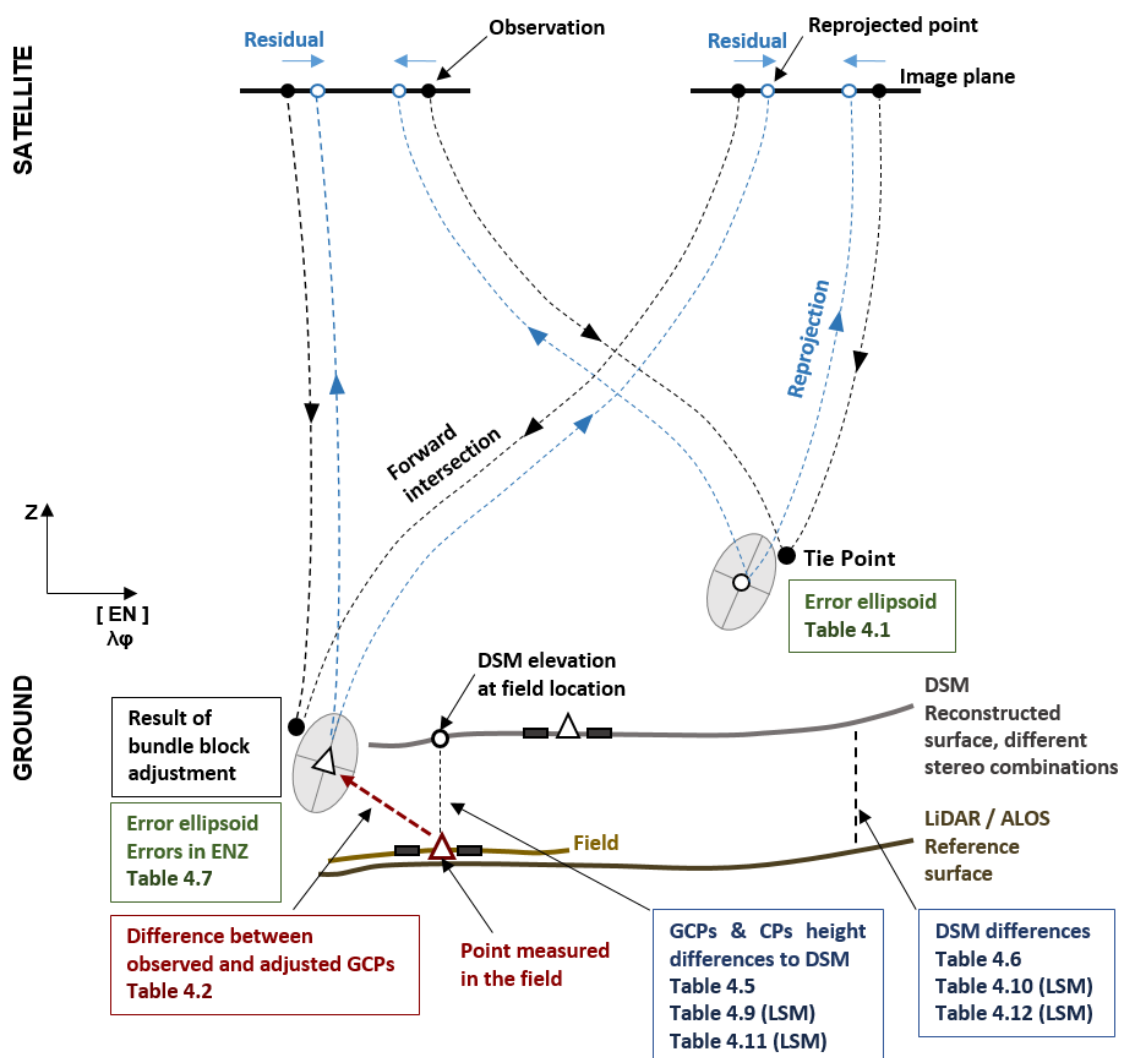

Figure 4.1: Accuracy measures for satellite image orientation and satellite-based DSMs with respect to GGPs, reference lidar DTM or ALOS DSM.

Graph in Figure 4.1 gives a comprehensive overview of the accuracy evaluation for both satellite image orientation and photogrammetrically derived DSMs with respect

to GCPs, reference lidar DTM (where available) or ALOS DSM. The main contributions results regarding the improvement of satellite-based DSMs and orthophotos are described in sections 4.3 and 4.4, while the 3D point cloud classification, small object height estimation and satellite image visibility analysis are given in sections 4.5, 4.6, and 4.7, respectively.

4.1   Sensor models parameter estimation

The VHR Pléiades and WorldView-3 satellite imagery for each of the four study areas were acquired in stereo or tri-stereo format, fact that made possible the extraction of 3D information by applying photogrammetric techniques. For accurate processing and to remove systematic errors, the image orientations provided through the RPCs were improved by estimating an affine transformation. The estimation was performed for each stereo/tri-stereo dataset in block adjustments, by using GCPs with known 3D coordinates (where available). As a result, to achieve sub-pixel accuracy in 3D geolocation, a set of correction parameters were determined and applied to each satellite image.

4.1.1   Satellite image orientation

Following the outline of the photogrammetric workflow described in Section 2.1, the satellite image orientation was performed for each scene individually. As a first step and where available, the GCPs were identified and measured in each image to stabilize the block and improve image orientations. Depending on the block configuration and overlap situations, the software allows the computation of different correction parameters: two shifts, two scales and two shears (on X and Y). Parameters for scale, which include the radial ephemeris and interior orientation errors (such as focal length and lens distortion errors) are computed for large images only (for more than 25 000 pixels) (Trimble 2016a). For example, in case of larger image extends in one direction (Allentsteig 01 - Pléiades and Allentsteig - WorldView-3), a single corresponding scale parameter is computed (Table 4.1). The highest shift values of -23.48 m in X and 40.14 m in Y direction are obtained for Barcelona cityscape. A detailed investigation is presented in Section 4.1.2.

During the adjustment of the stereo/tri-stereo scenes, TPs are automatically extracted in each satellite image. These are homogenously distributed in the entire scene with standard deviations of the adjusted coordinates ranging between 0.08 m and 2.12 m. The highest standard deviations are recorded for the elevations in the Himalaya datasets (1.57 m, 2.12 m, and 1.29 m; no GCPs were used in this computation). The minimum values in planimetry correspond to the Allentsteig WorldView-3 dataset (0.09 m in latitude and 0.08 m in longitude), while the smallest values in elevation correspond to the urban areas: 0.53 m and 0.38 m standard deviations for Barcelona and Stockholm, respectively. For all datasets, the standard deviations of the image block-orientation is around 0.5-0.6 pixels. When looking at the statistics for image residuals of automatic TPs (last six columns in Table 4.1), generally the values in cross-track direction (x-residual) are larger than those in-track direction (y-residual). When comparing the TPs RMSE y-residual within Allentsteig image triplet (0.14 pixels, 0.22 pixels, and 0.12 pixels for Forward, Nadir, and Backward, respectively), a larger value for the Nadir image in comparison to

Forward and Backward scenes is obtained. The RMSE of y-residuals were additionally checked for WorldView-3 Allentsteig (0.16 pixels, 0.21 pixels, 0.14 pixels for Forward, Nadir, Backward, respectively) and Pléiades Himalaya 01 triplets (0.25 pixels, 0.52 pixels, 0.23 pixels for Forward, Nadir, Backward, respectively), where the larger values corresponding to Nadir image indicate a similar behavior as for Allentsteig. Table 4.1 gives an overview of the satellite image orientation results for each study site.

The numerical accuracy of the estimated orientation was assessed with respect to the available GCPs and CPs for the imagery datasets, by computing the differences between the adjusted coordinates and the input ones. The root mean square error values of the 3D coordinates in units of meters and pixels are given in Table 4.2. For both sensors the RMSE values of the GCPs for latitude, longitude and elevation are at sub-pixel level, in the range of 0.30 to 0.60 pixels, showing that the block adjustment provides good results for subsequent processing. The only exemption, reaching a value of 0.36 m (1.20 pixels) in elevation corresponds to the Barcelona dataset. The small planimetric discrepancies suggest a GCPs identification in images better than 1/2 a pixel. As expected, higher accuracy is produced in cross-track direction (longitude). The larger statistics in along-track direction reflect the possible existence of residual systematic errors.

The Pléiades elevation accuracy (0.35 m for CPs in Allentsteig) is better than the result in (Bernard et al. 2012) of 0.49 m, and more favorable than the results in (Perko, Raggam, Gutjahr, et al. 2014), where 1 m RMSE in elevation were reported. Overall, the residual errors for Pléiades images at GCPs and CPs are comparable with the results obtained by Stumpf et al. (2014) reporting RMSEs of 0.44 m, 0.60 m and 0.50 m at 40 GCPs for latitude, longitude and elevation, respectively. For WorldView-3, the vertical accuracy fits to the results in F. Hu et al. (2016) which reported an elevation bias of less than 0.5 m for 6001 ground lidar CPs.

Table 4.1: Satellite imagery triangulation results. (*Image residual statistics shown for Forward image only, except Allentsteig 01.)

Study area Sensor type	No. of GCPs	No. of TPs/Image	Image transformations Shift in X/Y	Scale in X/Y	Sigma (px)	TPs Standard Deviation X (m) min	max	Y (m) min	max	Z (m) min	max	Image residuals for TPs observations x-residual (px) μ	σ	RMS	y-residual (px) μ	σ	RMS
Allentsteig 01 Pléiades																	
Forward	43	348	10.53 m X -7.53 m Y	3.56·10⁻⁶ Y	0.28	0.07	0.15	0.07	0.18	0.27	1.02	0.00	0.26	0.26	0.00	0.13	0.13
Nadir		341	10.32 m X -11.19 m Y	1.57·10⁻⁶ Y								0.00	0.28	0.27	0.00	0.21	0.22
Backward		345	10.54 m X -9.41 m Y	9.20·10⁻⁶ Y								0.00	0.26	0.26	0.00	0.12	0.12
Allentsteig 02 Pléiades	40	600/610/549	7.12 m X -7.41 m Y	1.63·10⁻⁶ X 1.41·10⁻⁵ Y	0.32	0.08	0.14	0.08	0.13	0.33	0.24	-0.01	0.24	0.25	-0.01	0.20	0.20
Allentsteig WorldView-3	36	574/603/572	-1.86 m X 1.65 m Y	1.36·10⁻⁷ X	0.31	0.05	0.12	0.04	0.11	0.21	0.59	-0.01	0.39	0.39	0.00	0.16	0.16
Barcelona WorldView-3	23	304/305	-23.48 m X 40.14 m Y	8.12·10⁻⁶ X 3.11·10⁻⁷ Y	0.33	0.15	0.25	0.14	0.24	0.36	0.53	0.00	0.33	0.35	-0.00	0.13	0.13
Stockholm WorldView-3	-	309/308	3.50 m X 1.57 m Y	-	0.37	0.17	0.22	0.11	0.14	0.31	0.38	-0.00	0.34	0.35	-0.00	0.14	0.15
Himalaya 01 Pléiades	-	402/439/400	1.00 m X 8.24 m Y	-	0.42	0.16	0.24	0.16	0.31	0.79	1.57	0.00	0.41	0.41	0.00	0.25	0.25
Himalaya 02 Pléiades	-	217/302/286	6.36 m X 11.24 m Y	-	0.43	0.21	0.61	0.13	0.23	0.77	2.12	-0.03	0.52	0.53	-0.00	0.18	0.17
Himalaya 03 Pléiades	-	268/306/216	0.60 m X 0.56 m Y	-	0.39	0.14	0.21	0.14	0.26	0.66	1.29	-0.01	0.38	0.39	0.00	0.16	0.16

Table 4.2: Root mean square error (RMSE) values for GCPs and CPs after block adjustment.

Site / Sensor	No. of GCPs/CPs	RMSE Values of Ground Checkpoint discrepancies (Units are Meters and Pixels) Latitude (along track)	Longitude (across track)	Elevation
Allentsteig 01 Pléiades	43 GCPs	0.19 (0.27)	0.17 (0.24)	0.27 (0.39)
	50 CPs	0.32 (0.46)	0.21 (0.30)	0.35 (0.50)
Allentsteig 02 - Pléiades	40 GCPs	0.35 (0.50)	0.31 (0.44)	0.34 (0.49)
Allentsteig WorldView-3	36 GCPs	0.14 (0.47)	0.13 (0.43)	0.11 (0.37)
	50 CPs	0.13 (0.43)	0.12 (0.40)	0.18 (0.60)
Barcelona – WorldView-3	23 GCPs	0.15 (0.50)	0.14 (0.47)	0.36 (1.20)

4.1.2 Accuracy of map projected satellite images

In the case of the Barcelona test site, in an initial stage, the planimetric geo-location accuracy of the provided images was visually checked against the available free OSM database. As map projected products at Stereo OR2A processing level, the images have an approximate geoposition, with planimetric shifts that vary within the area. Depending on the viewing angles for the stereo pair, the shifts may have different orientations. For the Barcelona test site, they are oriented to north-east for the first image and to south for the second image (Figure 4.2). This fact is proven by the resulted statistics for the 23 GCPs, as well (Table 4.3). For the forward image, when compared to the reference coordinates, the resulted mean values of the differences in X and Y directions (of 10.44 m and 6.87 m, respectively) are both positive, indicating a north-east shift. On the other hand, for the backward image, the mean of 3.85 m suggests a positive offset in X direction, while the negative mean of -12.64 m indicates a shift to south. The mean overall planimetric offsets are of 19 m and 23 m for the forward and backward images, respectively. The highest difference (of 46.5 m) was computed for a GCP located in the northern part, in the forested area, having the highest elevation of 211 m. This explains the relief displacements caused by the perspective projection, since the Barcelona images at OR2A processing level are map projected products to a constant elevation plane of 132.58 m. This effect is in detail described in Section 4.4.3.
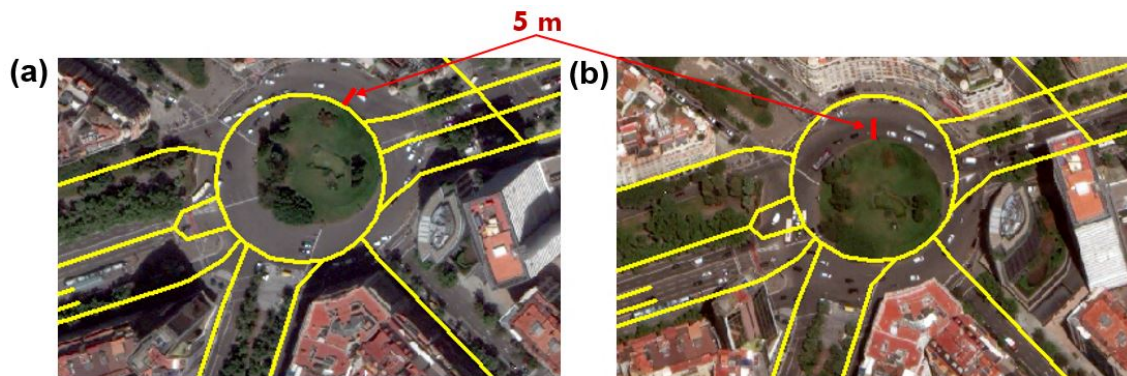

Figure 4.2: Planimetric shifts between reference OSM street lines (yellow) and the two stereo images for Barcelona corresponding to a traffic roundabout.

Table 4.3: Evaluation on 23 GCPs of the planimetric accuracy for the OR2A imagery over Barcelona (given values are in meters).

Barcelona Scenes (OR2A)	Planimetric displacement						Overall displacement		
	in X-direction			in Y-direction			XY		
	μ	σ	RMS	μ	σ	RMS	μ	σ	RMS
Forward	10.44	11.19	15.12	6.87	13.59	13.57	19.10	7.10	20.32
Backward	3.85	4.44	5.80	-12.64	24.13	24.42	23.00	10.29	25.10

To improve their geolocation accuracy and for removing the effects of image perspective (tilt), the satellite images were in a first step oriented by using the provided RPCs, which were refined with the available 23 GCPs and 304 TPs. In a second step, the images were orthorectified using the corresponding elevation model. The workflow for orthorectification is described in Section 4.4.

Although only computed for one scene, this gives an indication of the accuracy achieved without GCPs and without a DEM for orthophoto computation. The result is dominated by a systematic shift (here of 20 m) which can be reduced to below 1 m using GCPs (Table 4.2).

4.2 Accuracy analysis of DSMs from VHR satellite imagery

4.2.1 3D point cloud reconstruction

3D information extraction from satellite imagery was possible by applying the dense image matching algorithm for each site individually and through forward intersections, the three-dimensional positions of the points in object space (i.e. X, Y, Z) are determined. The final output are dense photogrammetric point clouds in LAS file format containing not only the 3D coordinates, but also the reflectance information from the three spectral bands (Red, Green and Blue). With a regular distribution (one point per each image pixel), the reconstructed point clouds have densities of 4 points/m^2 and of 12 points/m^2 for Pléiades and WorldView-3, respectively.

The photogrammetric point clouds describe the terrain surface with different land covers (bare soil, agricultural-, grass-lands) and the upper surface of the objects on it (vegetation, forested areas, infrastructure, buildings). The diversity of land cover and relief characteristics of the four datasets with their unique characteristics gives a comprehensive and detailed overview of the performance of 3D reconstruction from satellite imagery.

Generally, the resulted photogrammetric point clouds are characterized by a 3D geometry with smoothing effects in areas with a rapid change elevation (especially met at buildings roofs edges that look bevelled), occlusions (appear in urban areas, due to high buildings) and missing elevation information for small individual objects. Usually, the bigger differences occur in a buffer zone around buildings, because the extracted 3D information is not able to reconstruct the shape of vertical walls and tend to connect ground and buildings roof levels with slopes, generating a smoothing effect. Some detail views of the reconstructed point clouds from Pléiades and WorldView-3 over the investigated study sites are shown in Figure 4.3.

Generally, fields, grasslands and agricultural areas met in the Allentsteig dataset, have a good appearance with regular distributed points in the photogrammetric point clouds, due to the smooth, open surfaces, which are visible in all the three images of the acquisition. As expected, smoothing effects are visible at buildings roof edges. In contrast to Pléiades, the WorldView-3 point cloud shows a better preservation of information and object details on the terrain surface (Figure 4.4). This can be explained by the better spatial resolution (0.31 m GSD) and the acquisition geometry with higher values for the viewing angles and a larger convergence angle on the ground (25°). All these factors allow a better 3D reconstruction resulting in a denser point cloud with reduced smoothing effect compared with Pléiades point clouds. In the case of Allentsteig dataset, very few small regions were not reconstructed due to occlusions. Nevertheless, these no data areas were filled in a following step, where interpolation algorithms were employed.

On the other hand, urban environments present specific challenges and are the most problematic for the reconstruction process, due to their complex morphology of the urban texture. The two investigated urban test sites of Barcelona and Stockholm
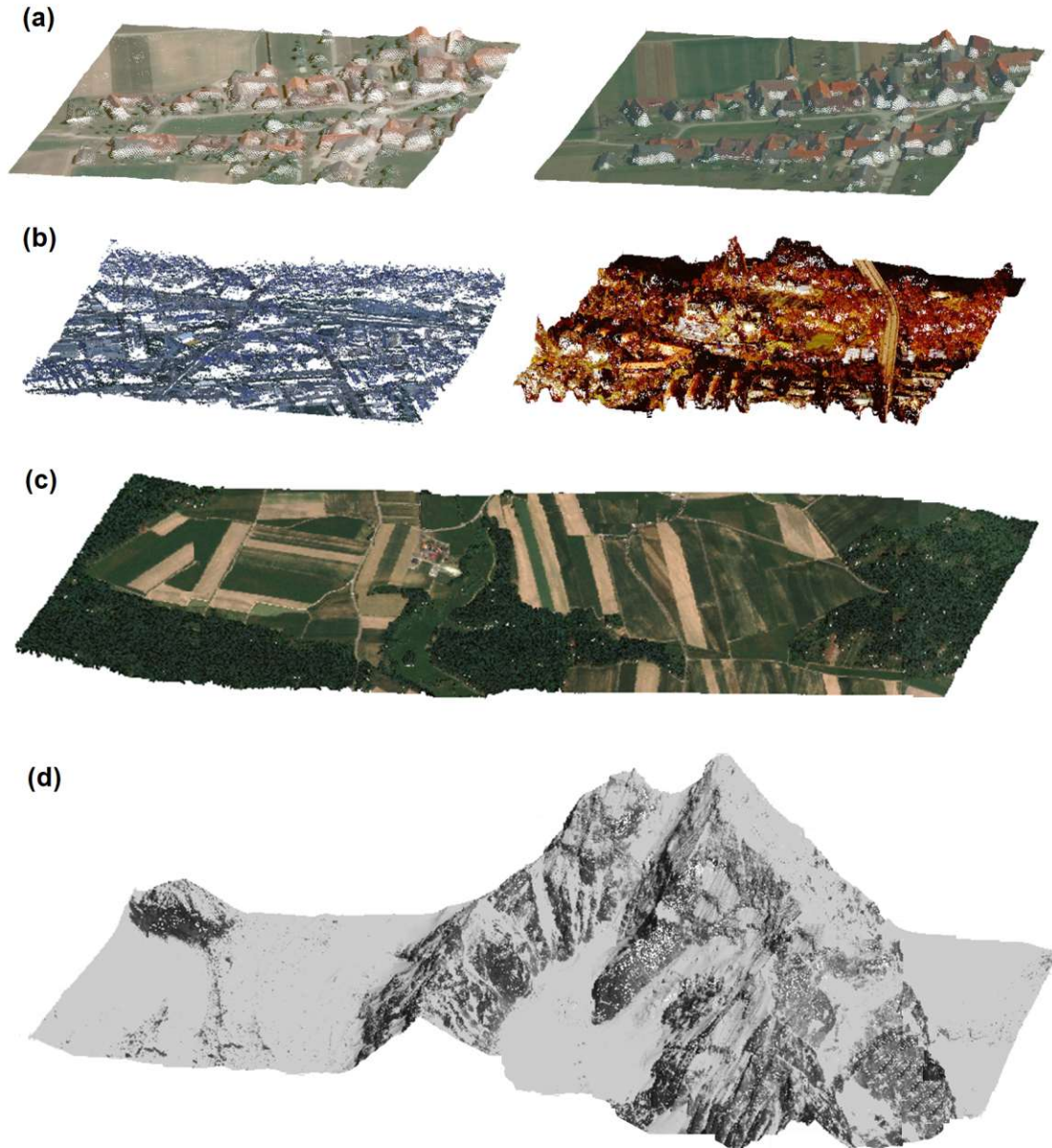
Figure 4.3: Photogrammetric 3D point clouds from VHR Pléiades and WorldView-3 satellite imagery over the study sites - detail views: (a) rural area in Allentsteig with Pléiades (left) and WorldView-3 (right); (b) urban areas of Barcelona (left) and Stockholm (right) with WorldView-3; (c) field and forested area in Allentsteig (Pléiades); (d) Himalaya mountains (Pléiades).

describe the extremely dense urban centers, with very narrow roads and buildings of varied sizes, shapes and heights, but also open spaces (parks), river channels and water surfaces (Figure 4.3(b)). From a visual inspection of the reconstructed point clouds, severe discrepancies frequently occur in narrow roads and urban canyons (up to several meters), due to the complex morphology, acquisition geometry and specific urban texture. A careful observation of the reconstructed point clouds shows that buildings that have large roof area with bright colors tend to give better representation. In contrast, high-rise and small buildings are not reconstructed appropriately. Moreover, by using stereo acquisitions with only two images, many regions were not
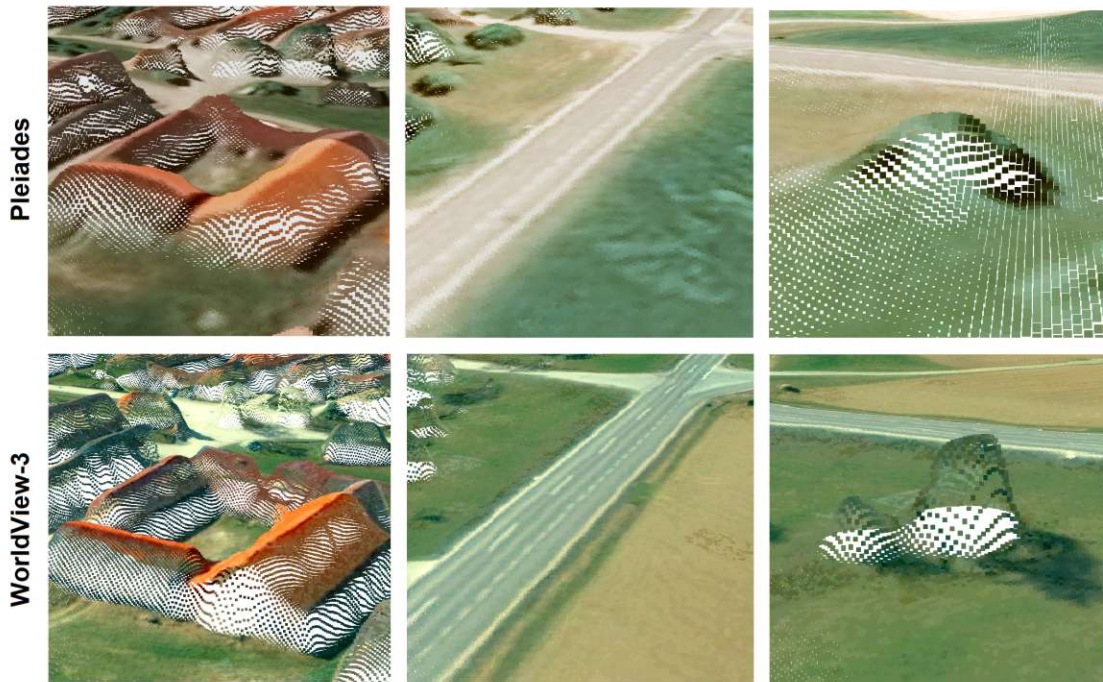
Figure 4.4: Comparative examples of Pléiades (first row) and WorldView-3 (second row) point clouds for buildings (first column), roads (second column) and trees (third column).

reconstructed due to occlusions (large elevation difference between buildings/trees and surrounding ground).

The more problematic parts are the narrow streets, where elevations are always overestimated. Detail views of the point clouds in the dense urban centers with narrow streets are shown in Figure 4.5. The analysis of the two profiles in Barcelona city center makes clear the inability of the reconstructed point clouds in describing correctly street geometry and, more generally, vertical facades of buildings, resulting in a local overestimation of the ground level. The narrow streets of approximately 10 m width between high buildings clearly have an elevation overestimation between 5 and 25 m. While ground surface and first building in profile (1) (from left to right) have quite good reconstructed elevations, the second building seems to have an overestimation with approximately 4 m in the middle roof parts, due to the roof structure with higher elevations at edges (67 m) compared to the middle (63 m). The intern courtyard on the right has overestimated elevations with irregular satellite-based point distributions. Differences occur also in a buffer zone around buildings, because the extracted 3D information is not able to reconstruct the shape of vertical walls. For Barcelona dataset and these profiles, the available lidar data with lower point density of less than 1 point/m^2 was used. Since no reference GCPs were available for improving the satellite image orientations, an elevation offset of approximately 25 m above sea level is visible in the Stockholm dataset (Figure 4.5(b)). At a first impression, the reconstructed water surface close to the coast seems to have a smooth appearance, but actually waves of more than 5 m and low peaks are visible.

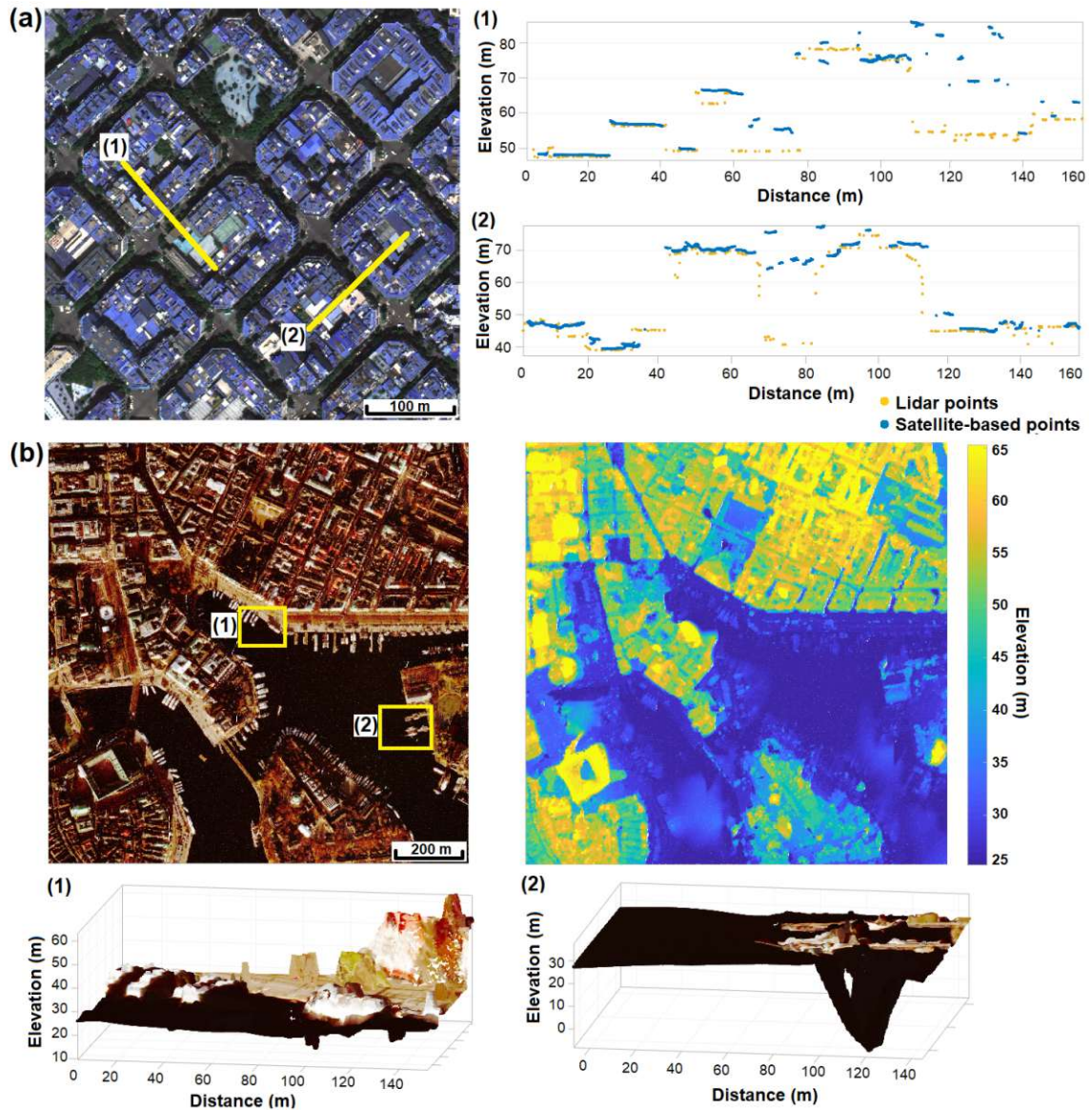In the case of Himalaya test site, the Pléiades tri-stereo panchromatic images were

Figure 4.5: Photogrammetric 3D point clouds for urban areas - detail views: (a) Barcelona with two profile sections (a buffer of 1 m was considered for point selection); (b) Stockholm RGB- and elevation- colored point clouds with zoom-in view of the boats in port (areas marked with yellow rectangles).

employed in the photogrammetric process. Visually, the resulted 3D point clouds from spring 2015 have a nice appearance, with details for the mountain peaks covered by snow, cliffs, rocks, and valleys. Void areas are present especially at very steep slopes. Figure 4.6 shows the 3D view of the reconstructed Everest and Cho Oyu mountain peaks. Scattering, noisy effects and outliers from dense image matching are visible at the eastern side slope of Everest peak in Spring 2015 acquisition, due to poor image texture with white snow and ice, which affects dense image matching quality. On the other hand, the reconstruction from the Autumn 2015 acquisition shows better results.

The low image texture affects also the automatic extraction of TPs in images (Figure 4.7). While Spring acquisition is characterized by a lower density of TPs due to the
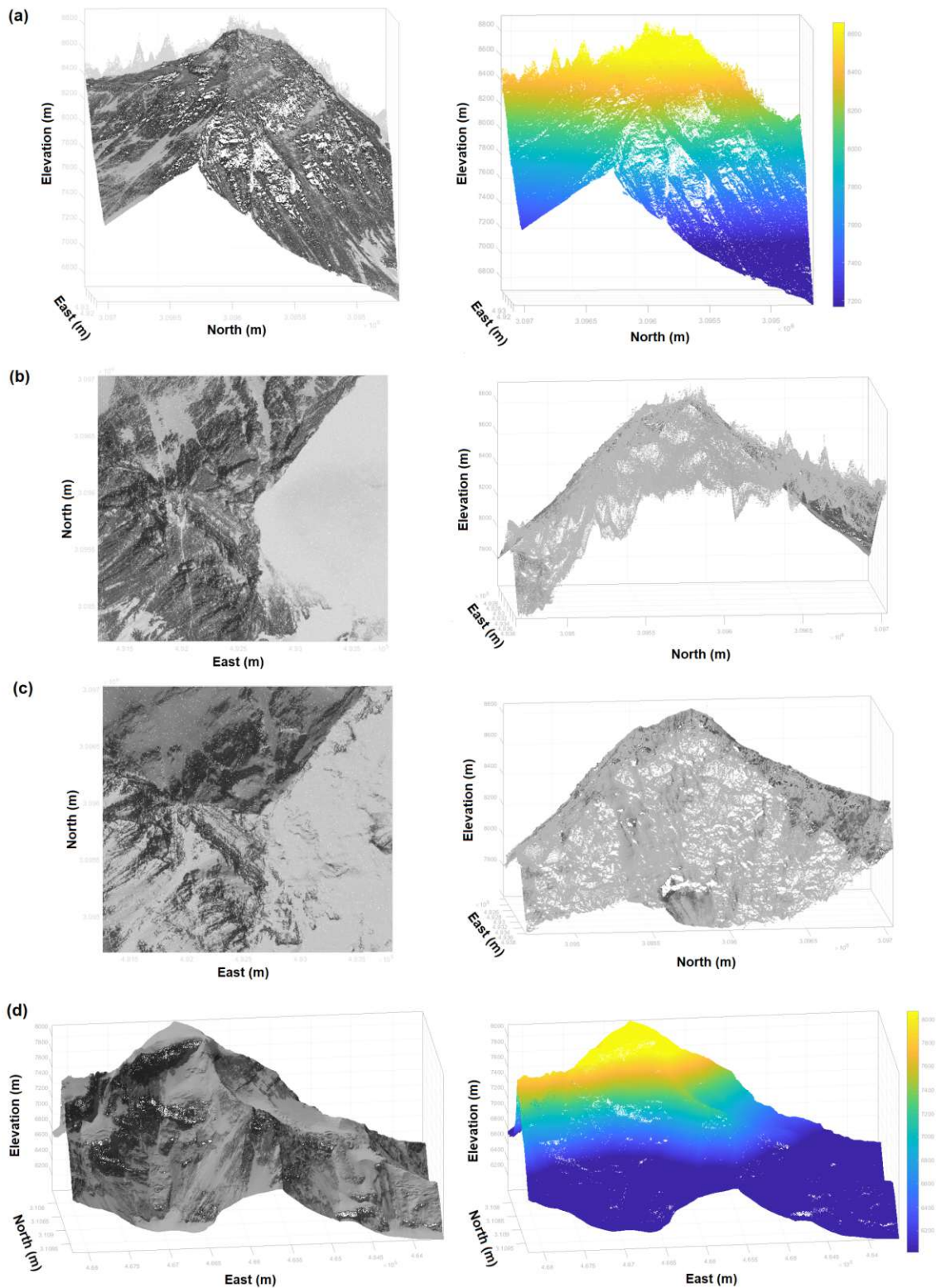
Figure 4.6: Photogrammetric 3D point clouds over Himalaya mountain peaks: (a) Everest - panchromatic, colored by elevation (Spring 2015 acquisition); (b) Everest - top view, and eastern mountain side (Spring 2015 acquisition); (c) Everest - top view, and eastern mountain side (Autumn 2015 acquisition); (d) Cho Oyu - panchromatic and colored by elevation.

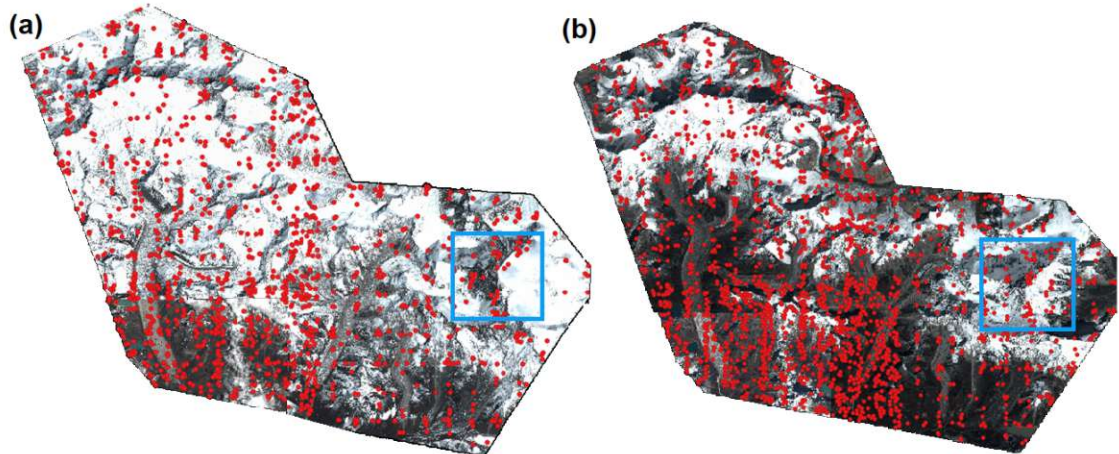wider snow coverage area, the Autumn dataset contains a larger number of TPs.

Figure 4.7: TPs distribution in Himalaya images with marked location of Everest peak surrounding: (a) Spring 2015 acquisition; (b) Autumn 2015 acquisition.

Table 4.4: Dense image matching results for each study site and image combination.

Scene Comb.	Matching Time (h:min)	LAS file (GB)	No. of pts (mil.)	Scene Comb.	Matching Time (h:min)	LAS file (GB)	No. of pts (mil.)
Allentsteig 01 (Pléiades)				**Allentsteig (WorldView-3)**			
FNB	10h 12'	15	667	**FNB**	33h 28'	27	1 049
FB	09h 15'	15	583	**FB**	21h 24'	26	1 048
FN	10h 35'	15	584	**FN**	24h 13'	26	1 048
NB	10h 23'	15	597	**NB**	24h 20'	26	1 048
Allentsteig 02 (Pléiades)				**Barcelona (WorldView-3)**			
FNB	14h 14'	35	1 332	**FB**	13h 10'	25	961
Stockholm (WorldView-3)				**Himalaya 01 (Pléiades)**			
FB	3h 04'	8	291	**FNB**	29h 40'	32	1 271
Himalaya 02 (Pléiades)				**Himalaya 03 (Pléiades)**			
FNB	29h 22'	31	1 328	**FNB**	29h 24'	33	1 425

Table 4.4 contains an overview of the results from dense image matching for each study site and scene combination. To investigate the impact of the satellite viewing geometry on the reconstruction quality performance, in the case of first two datasets (Allentsteig 01-Pléiades and Allentsteig-WorldView-3), besides the tri-stereo scene, FNB (Forward-Nadir-Backward), dense image matching was executed individually for each two-set pairwise image combination: FB (Forward-Backward), FN (Forward-Nadir), and NB (Nadir-backward) (Section 4.2.4). Pléiades FNB triplet provides the largest point cloud (667 milion points), compared with the three stereo pairs having 583 (FB), 584 (FN) and 597 (NB) milion points. Approximately same number of 3D points are generated for all four Allentsteig (WorldView-3) combinations (1048 milion points). The number of reconstructed points for Himalaya increases from 1 271 mil. in Spring 2015 to 1 425 mil. in Spring 2016.

All image matching processes were performed using the same machine with a core of 3.50 GHz and 32 GB RAM. Depending on the covered area and image resolution, the minimum and maximum computation time for generating the 3D point clouds are 3

and 33 hours corresponding to Stockholm (27 km^2) and Allentsteig (WorldView-3, high point density) datasets, respectively.

4.2.2 DSM derivation

The resulted photogrammetric point clouds are used as input for DSM derivation. Usually the transition from 3D points to 2.5D grid models is done through interpolation that tends to smooth the input elevation values. The accuracy of obtained models highly depends on the adopted interpolation strategy and the choice of grid resolution. For the current study areas, a robust moving planes interpolation was used, that estimates a best fitting tilted plane for every point considering a search radius of maximum 1 m. This interpolation strategy is actually a compromise between achieving fidelity to the true surface and respecting the limitations according to the density and accuracy of the data source. For minimizing the smoothness effect, a small size of the grid cell was selected (0.5 x 0.5 m) and a relatively small neighbourhood definition for the interpolation. To determine the direct degree of smoothing, a raster containing the maximum elevations in each 0.5 x 0.5 m cell was computed. Using this as a reference, the 2.5D Pléiades interpolated model showed a RMSE value of 0.079 m, while the WorldView-3 DSM had 0.098 m. The latter was slightly higher, but these values (at sub-decimetre level) still showed that the smoothing effect of the interpolation will not significantly influence further analysis.
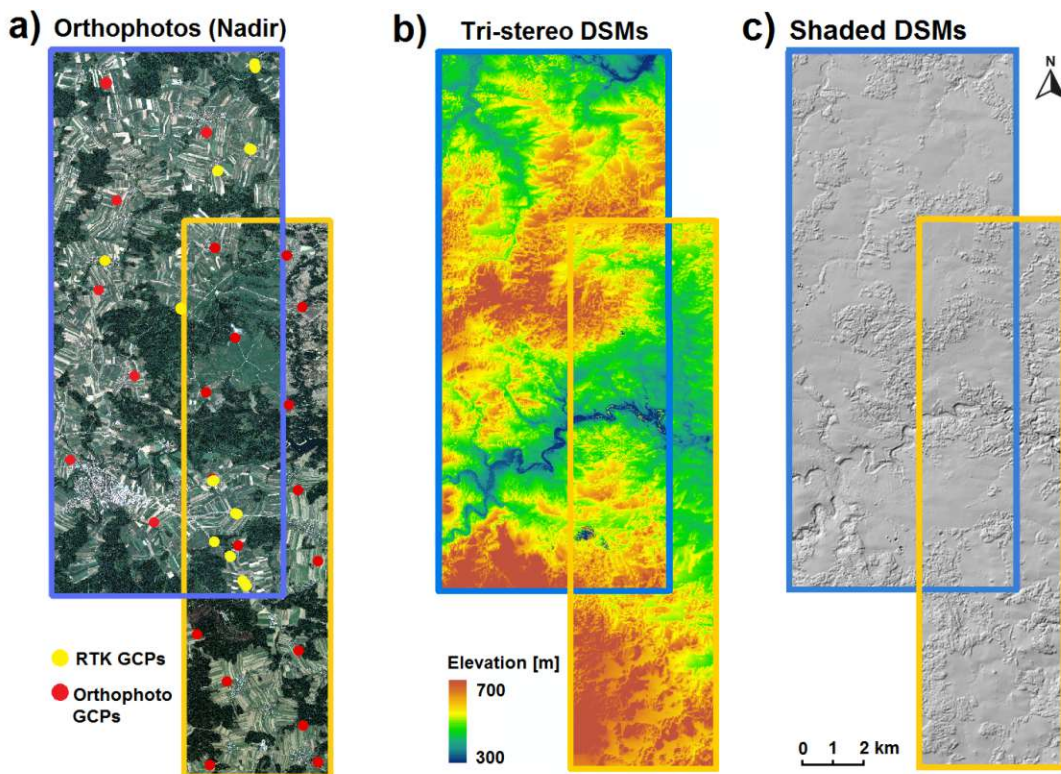

Figure 4.8: (a) Allentsteig orthophoto visualised as true color RGB with overlaid RTK GCPs (yellow circles) and manually measured GCPs (red circles); (b) elevation colour coded DSMs; (c) shaded reconstructed DSMs.

Based on the described interpolation strategy, DSMs over the four test sites were generated by using the scientific software OPALS (Orientation and Processing of

Airborne Laser Scanning Data) (Pfeifer et al. 2014). To analyze the full potential and limitations of 3D reconstruction from satellite imagery, the resulted photogrammetric point clouds were not edited nor filtered before performing the interpolation step. The diversity of land cover, relief and topography for each dataset, is clearly visible in both elevation-colored and shaded reconstructed DSMs.

Characterized mainly by agricultural fields, open and forested areas, the obtained DSMs for Allentsteig from both sensors (Pléiades and World-View-3) are generally good, with elevations ranging from 330 to 700 m (Figure 4.8). The only major exception appears at rivers and surfaces covered by water, where, as expected, problems in image matching occur. Detailed views of the DSMs in rural and forested areas are shown in Figure 4.9. Since lidar data is available for this region, additionally the nDSMs were derived by subtracting the reference lidar DTM from the Pléiades and WV-3 DEMs, respectively.
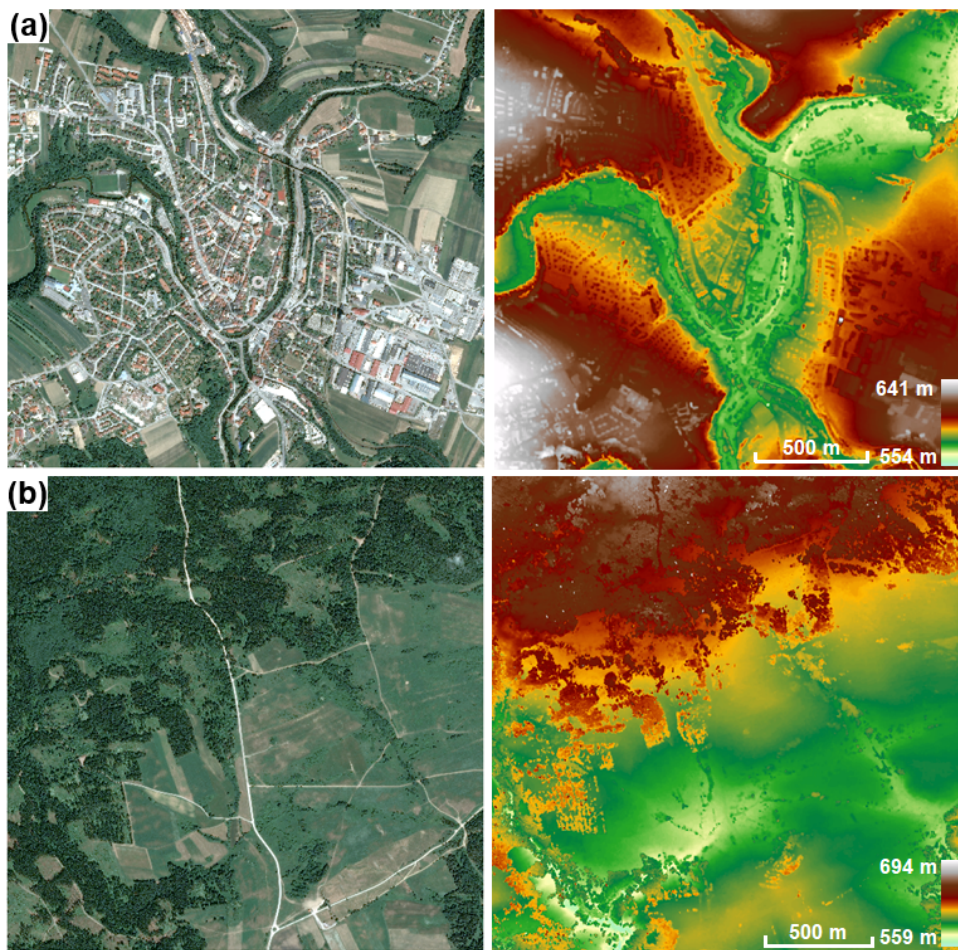

Figure 4.9: Detailed visualisations in Allentsteig orthophoto and DSM: (a) rural area (b) forests and open fields.

Considering the complex morphology of the urban areas of Barcelona and Stockholm, an extremely high level of detail is required to correctly describe the shape of buildings and infrastructures. For this reason, the DSM interpolation was performed at the highest possible resolution, which is actually the original resolution of the input satellite imagery (0.30 m). However, despite their high resolution, the reconstructed DSMs for Barcelona and Stockholm from stereo satellite imagery are
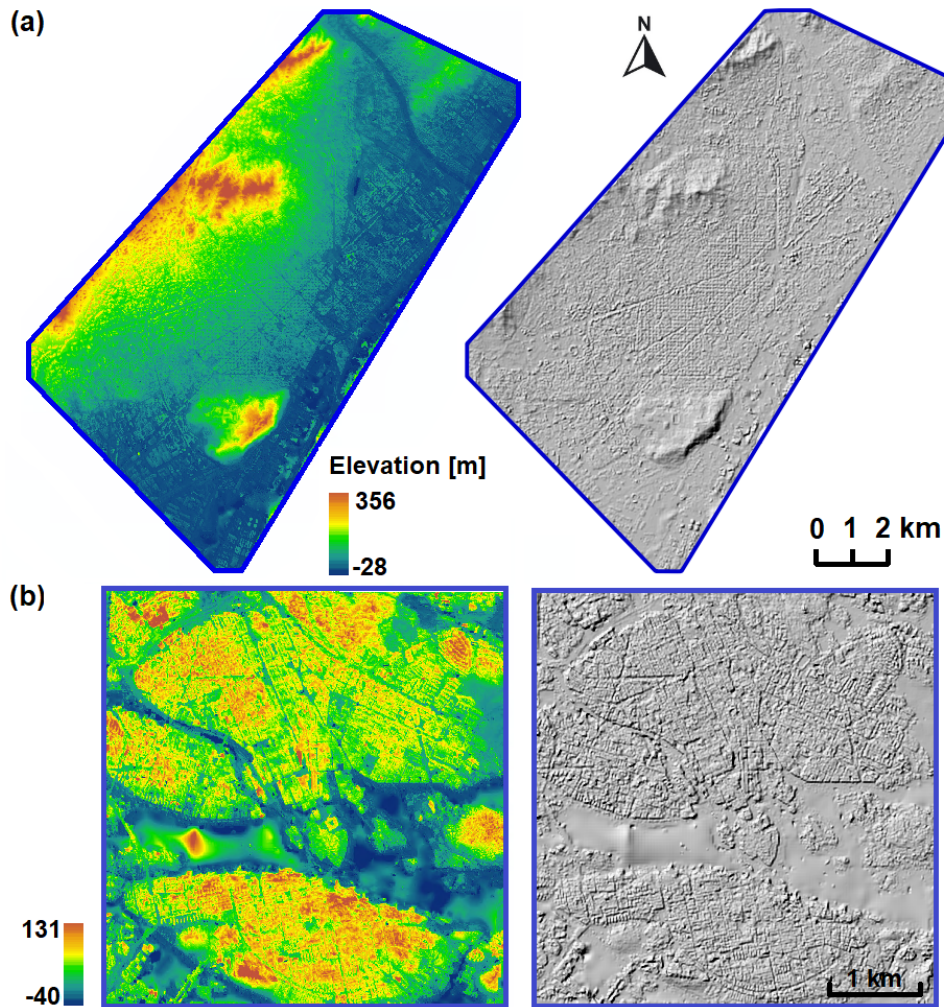
Figure 4.10: Digital Surface Models colored by elevation and shaded views for the urban areas: (a) Barcelona and (b) Stockholm.

characterized by a lower quality and completeness, with artefacts, noise and imperfections, resulting in a rough building shape representation (Figures 4.11 and 4.12). Moreover, the use of only two images (stereo acquisitions) with wide baselines in such complex scenes, leads to mismatches, missing values, and occlusion problems particularly at building edges (e.g. void areas, gaps). These were further filled in a post-processing step, by using the elevation information of the neighboring boundary pixels. The negative elevations of -28 and -40 m in Figure 4.10 show the limitation effects from dense image matching especially over water surfaces.

A careful look of the DSMs of Stockholm and Barcelona shows that buildings with a large roof area tend to give a good representation (Figure 4.11), while the high-rise and small buildings are not appropriately captured by the model (Figure 4.12(a)), as already observed in the 3D reconstructed point cloud. The example in Figure 4.12(b) shows a detail view for a street intersection and surrounding buildings in Barcelona with shadowing and occlusion effects. Even if buildings appear in the DSM with irregular contours they have quite well reconstructed heights and the elevation of the two intersecting streets is thoroughly constant. In contrast, moving, parked cars and areas surrounding buildings do not show any height information.
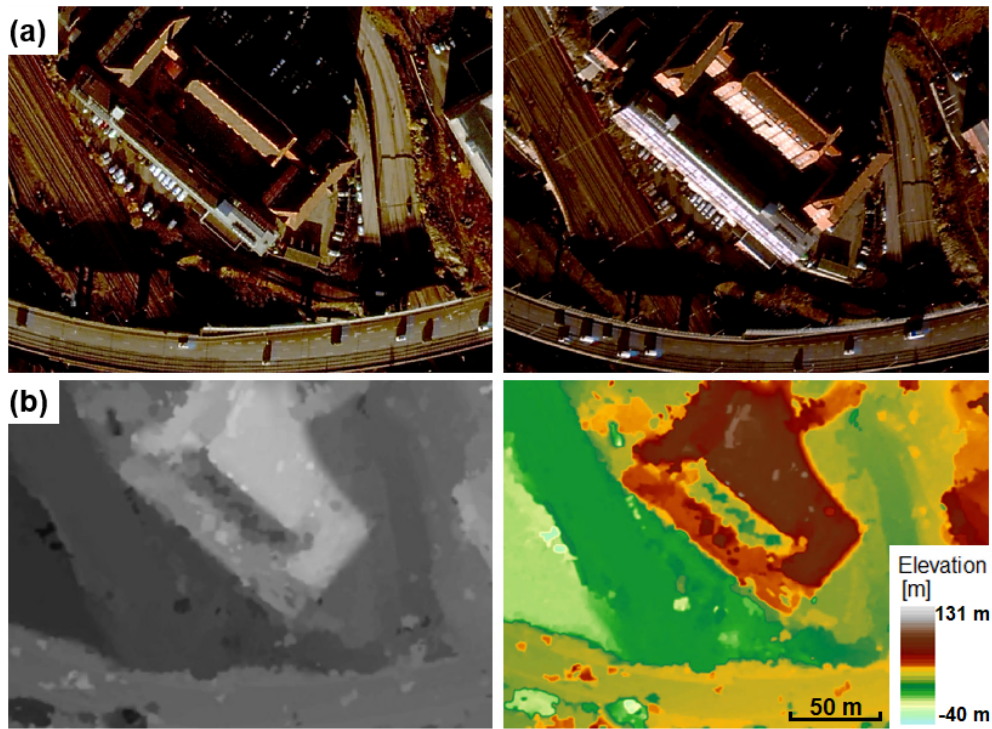
Figure 4.11: Detail view in Stockholm: (a) forward and backward satellite images; (b) surface model rendered in singleband gray and colored elevation.
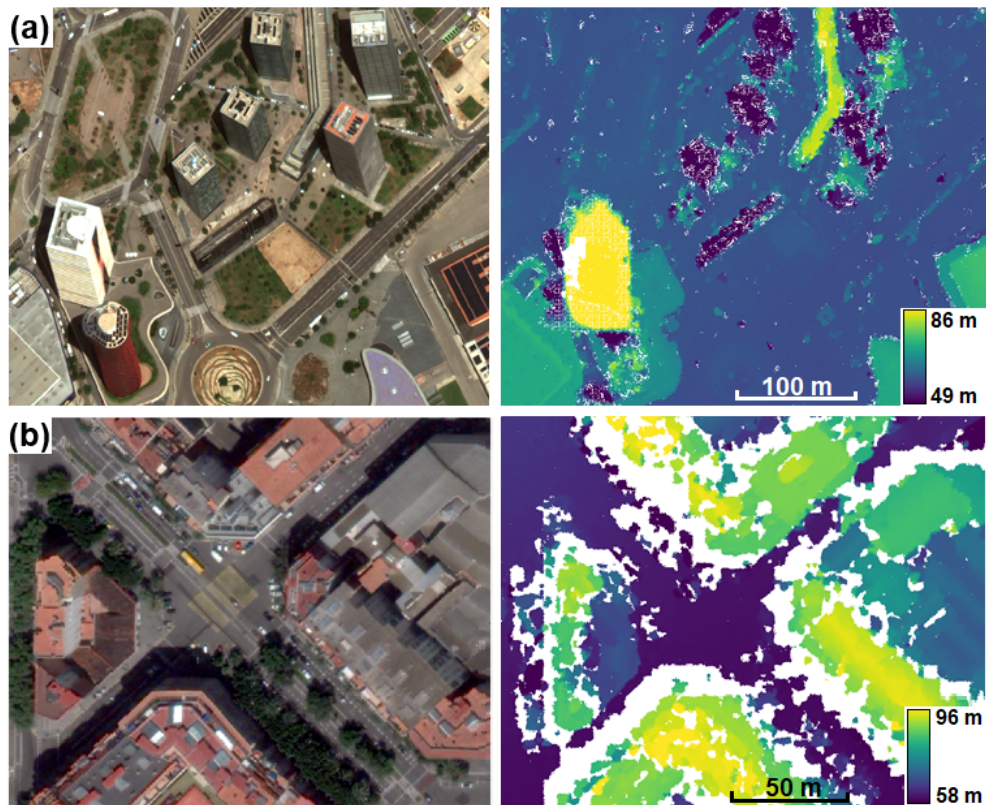

Figure 4.12: Bacelona zoom-in visualisations in the pansharpened WV-3 images and corresponding DSMs: (a) high-rise buildings; (b) street intersection, buildings, cars.

The computed Pléiades DSMs over the Himalaya mountain range show a good elevation reconstruction (between 4300 m and 8829 m) with visible glacier tongues, valleys, mountain crests, ridges and peaks (Figure 4.13). However, there are locations where image matching algorithm encountered difficulties in finding corresponding points, because of low image texture with shadows, white snow and ice covered surfaces.
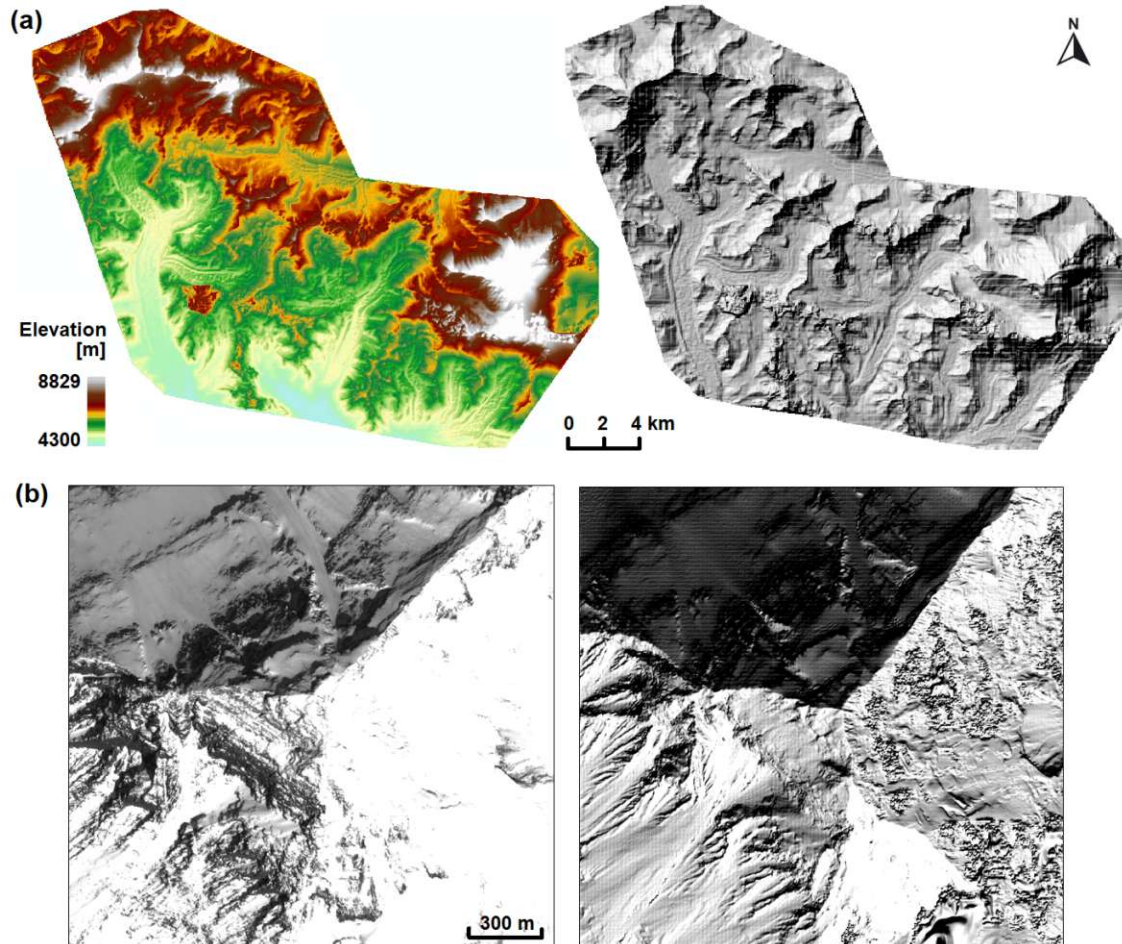

Figure 4.13: Overview of Himalaya DSM: (a) colored by elevation and shaded view for the entire available dataset; (b) detail-view of Everest peak in the pachromatic image (left) and shaded view of the reconstructed DSM (right).

4.2.3 Accuracy evaluation

The elevation accuracy of photogrammetrically derived DSMs for each study area was evaluated against GCPs, reference lidar DTM (when available) or ALOS DSM.

The results are summarized in Table 4.5, where the RMSEs in elevation are in the range of 1 pixel for Allentsteig (all cases) and Barcelona, when the provided RPCs were refined with the help of GCPs. Higher vertical offsets are obtained when no geolocation improvement is performed.

At a wider scale, the statistics of elevation differences between the satellite-driven DSMs and available reference models are given in Table 4.6. All three Allentsteig and Barcelona results were checked against lidar data (at 1 m and 2 m resolutions

Table 4.5: Elevation accuracy of photogrammetrically derived DSMs - statistics at available GCPs (* mountain peaks).

Site (Sensor)	No. GCPs	Mean (m)	Std (m)	RMSE (m)
Allentsteig 01 (Pleiades)	43	0.05	0.12	0.57
(no RPC bias correction)		16.50	0.32	16.01
Allentsteig 02 (Pleiades)	40	0.01	0.59	0.58
(no RPC bias correction)		16.52	0.82	17.14
Allentsteig (WorldView-3)	36	−0.02	0.31	0.31
(no RPC bias correction)		2.43	0.56	2.93
Barcelona (WorldView-3)	23	0.11	0.29	0.33
(no RPC bias correction)		3.28	0.95	3.58
Stockholm (WorldView-3)	-	-	-	-
Himalaya 01 (Pleiades)		0.03	10.94	10.13
Himalaya 02 (Pleiades)	*7	5.10	8.67	9.49
Himalaya 03 (Pleiades)		4.22	6.88	7.92
(no RPC bias correction)				

Table 4.6: Vertical accuracy assessment of Pléiades and World-View-3 DEMs against Lidar and ALOS elevation models (given values are in meters).

| Site	Sensor	No. of GCPs	Reference model (resolution)	Mean	Std	σMAD	RMSE
Allentsteig 01	Pléiades	43	Lidar (1 m)	0.80	0.53	0.51	0.96
		43	ALOS (30 m)	0.57	2.84	2.38	2.86
		-	Lidar (1 m)	17.79	1.21	0.59	17.83
Allentsteig 02	Pléiades	40	Lidar (1 m)	0.82	0.75	0.72	1.12
		-	Lidar (1 m)	18.10	2.02	0.75	18.32
Allentsteig	WV-3	36	Lidar (1 m)	-0.16	0.33	0.33	0.37
		-	Lidar (1 m)	3.76	0.55	0.26	3.80
Barcelona	WV-3	23	Lidar (2 m)	0.13	0.43	0.37	0.45
		-	Lidar (2 m)	4.21	0.76	0.54	4.25
Stockholm	WV-3	-	ALOS (30 m)	5.07	9.41	8.23	10.69
Himalaya 01	Pléiades	-	ALOS (30 m)	-1.26	23.20	17.76	23.24
Himalaya 02	Pléiades	-	ALOS (30 m)	3.95	21.91	9.34	22.26
Himalaya 03	Pléiades	-	ALOS (30 m)	4.32	24.16	16.51	24.76
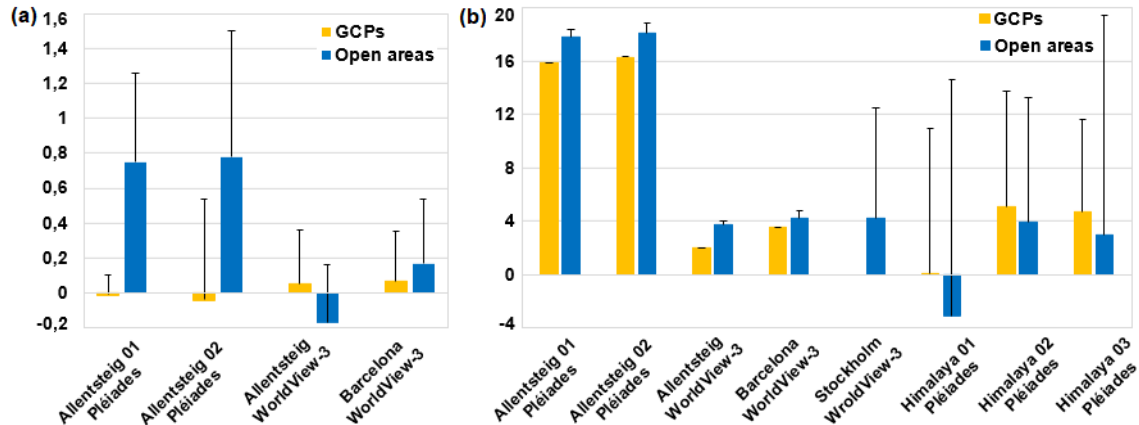
Figure 4.14: Statistics on elevation (median and robust standard deviations) of satellite-based DSMs at GCPs and in open areas: (a) when RPCs are bias-corrected with GCPs (with reference to lidar); (b) when RPCs are not bias-corrected (with reference to lidar and AW3D30).

for Allentsteig and Barcelona, respectively), while for Stockholm and Himalaya, the ALOS DSM (30 m resolution) was used. Similar to ground checkpoint results, RMSE values are less than 1.5 pixel, when RPCs are refined by using GCPs. Generally, when no RPC bias-correction model is applied, the satellite-driven DSMs show

systematic errors with positive Z offsets and higher RMS values. When compared to ALOS reference model, the results for Himalaya dataset are influenced not only by the time interval of more than 4 years between different acquisitions together with the seasonality (2015, 2016 spring, autumn for Himalaya and 2006 - 2011 for ALOS), but also by the lower ALOS resolution of 30 m.

A graphic overview of the statistic results for all datasets is given in Figure 4.14, where the median and robust standard deviations at GCPs and in open areas are shown for the approaches: with and without bias-corrected RPCs by using GCPs. In (a), the smaller median and robust standard deviations in open areas for the satellite WorldView-3 based DSMs indicate a higher accuracy of the World-View-3 models compared to Pléiades ones. Apparently with better values for the medians, but with larger standard deviations (graph (b) in Figure 4.14), the satellite-based DSMs for Stockholm and Himalaya are compared with AW3D30 model.
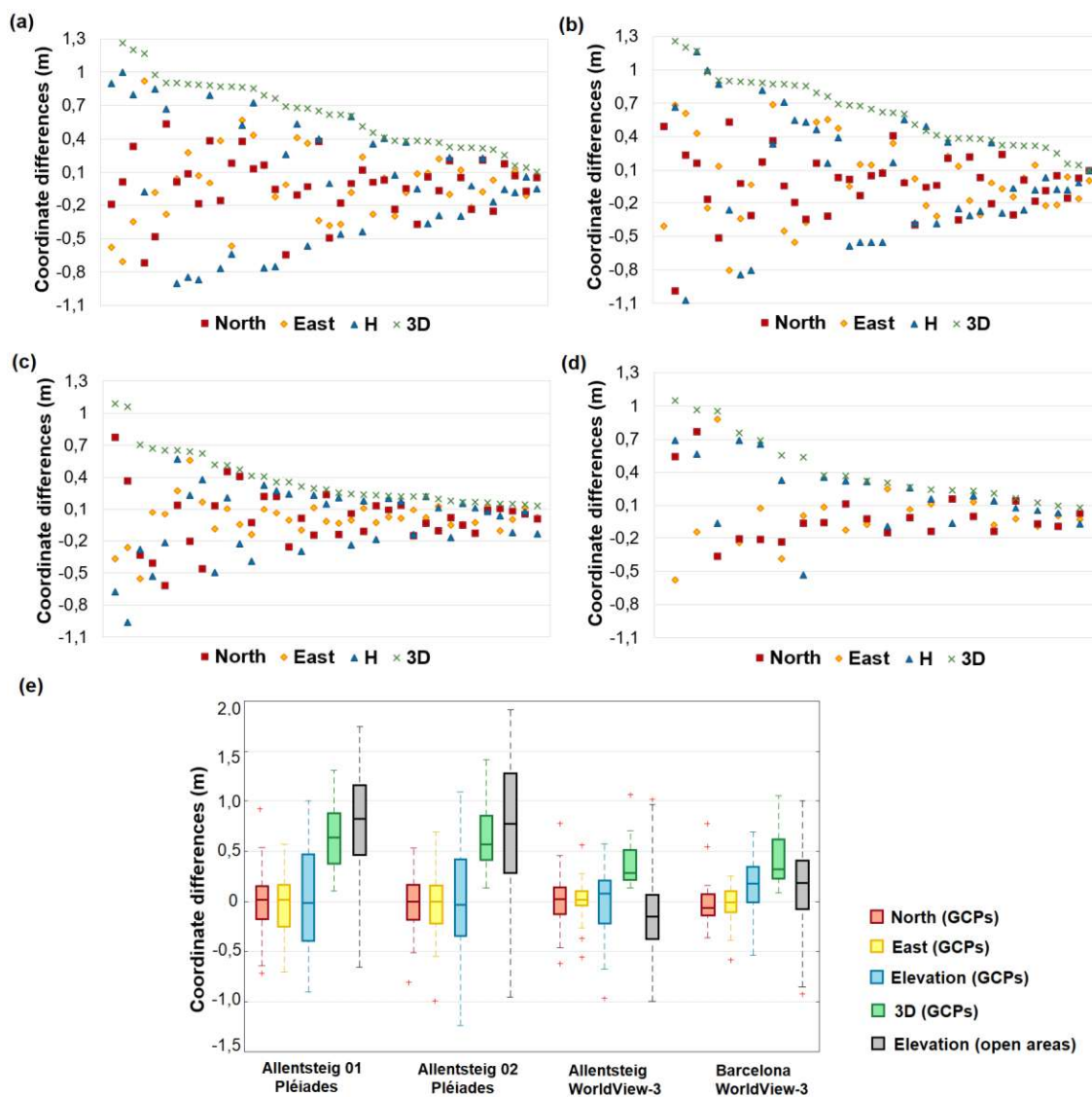

Figure 4.15: Comparisons between GCPs coordinates and stereoscopically measured on the epipolar images for: (a) Allentsteig 01 (Pléiades); (b) Allentsteig 02 (Pléiades); (c) Allentsteig (WorldView-3); (d) Barcelona (WorldView-3); (e) Statistics on GCPs coordinates and on elevation in open areas.

Table 4.7: Accuracy statistics at GCPs for Pléiades and World-View-3 DEMs over Allentsteig and Barcelona.

	East (m)				North (m)				Elevation (m)				3D (m)			
	μ	Med	σ	RMS	μ	Med	σ	RMS	μ	Med	σ	RMS	μ	Med	σ	RMS
A1(P)	-0.01	0.02	0.29	0.32	-0.02	0.01	0.22	0.27	0.05	-0.01	0.12	0.57	0.64	0.64	0.16	0.71
A2(P)	0.01	0.00	0.35	0.34	-0.03	0.00	0.29	0.29	0.01	-0.04	0.59	0.58	0.64	0.57	0.34	0.72
A(W)	0.01	0.00	0.17	0.17	0.02	0.01	0.26	0.26	-0.02	0.06	0.31	0.31	0.35	0.25	0.26	0.44
B(W)	-0.01	0.00	0.23	0.26	-0.01	-0.01	0.22	0.24	0.11	0.07	0.29	0.34	0.34	0.26	0.29	0.49

The accuracy of the computed satellite imagery aerial triangulation was evaluated in 3D by comparing the GCPs coordinates - measured directly on the oriented satellite images - with those measured on the terrain (or extracted from the orthophoto and lidar, after case). Statistic results for Allensteig and Barcelona study sites (where GCPs are available) are given in Table 4.7. Figure 4.15 shows the graphs of comparisons at GCPs, where the coordinate differences are less than 1 m for most of the points. As expected, values are a bit smaller for WorldView-3 than for Pléiades, but elevation differences are in the same order of magnitude as differences in East and North.

The planimetric geolocation accuracy of the Stockholm DSM was investigated in comparison with the free OSM street shape files (Figure 4.16).

Figure 4.16: DSM Stockholm geolocation: (a) orthophoto with marked area of interest and zoom-in in the DSM with OSM-reference and manually digitized street axes; (b) offet in planimetry.

For this, the axis of streets located inside a selected area of interest (2.12 km^2) were

manually digitized. The visual observable planimetric offsets to the OSM lines are between 0 to a maximum value of 5.5 m. Problems occur since some streets are not distinguishable in the DSM (but available in the OSM data); only 82% of the total number of streets in this area are identifyable in DSM. This happens due to the low sun angle (11.5°) at the acquisition time (17.11.2017), leading to dark satellite images.
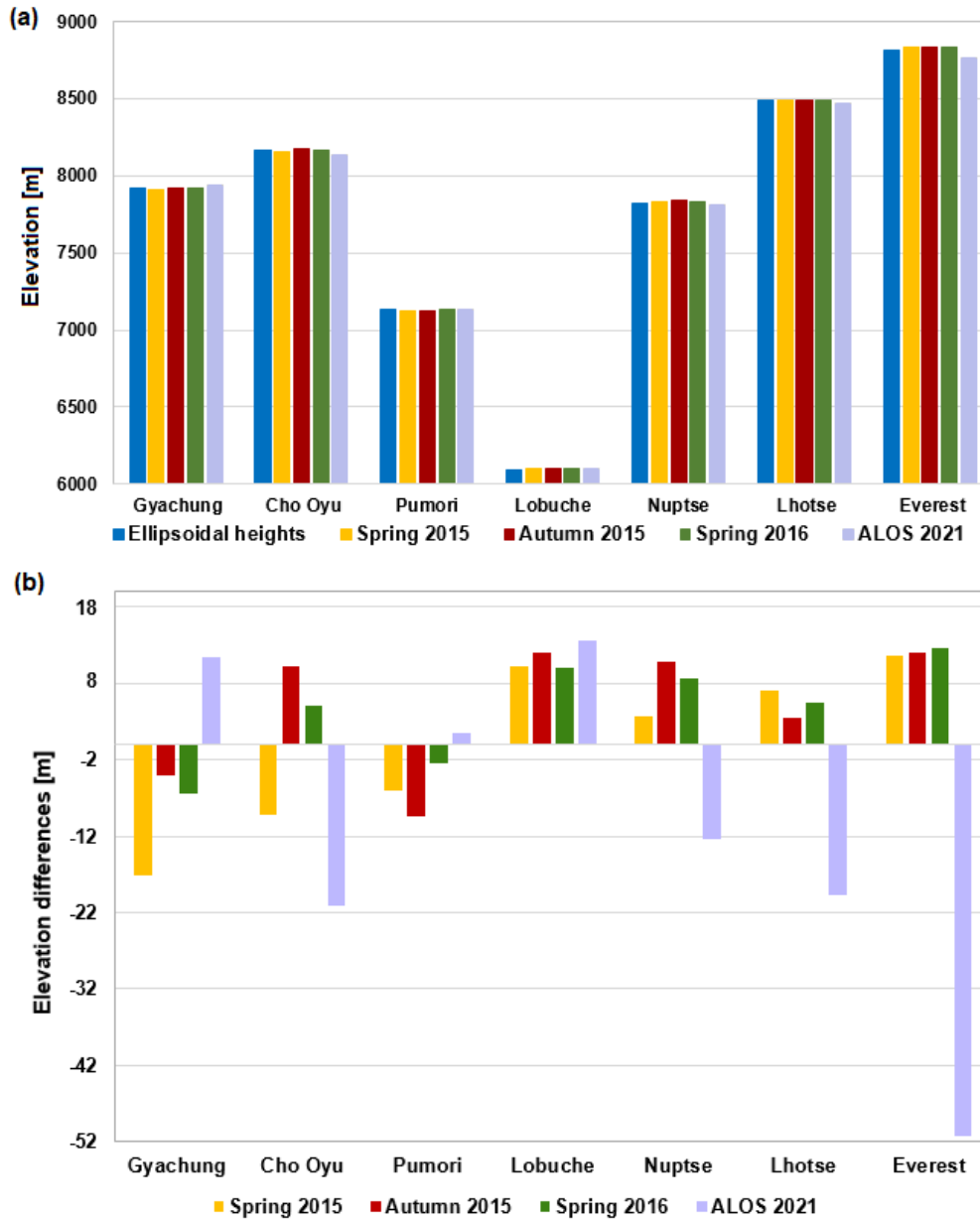

Figure 4.17: Highest peaks in Himalaya mountains: (a) Overall elevations; (b) Elevation differences with respect to the ellipsoidal heights.

In the case of Himalaya mountains site, the highest peaks were considered for the quality assessment. The reference heights are orthometric heights referred to the EGM96 geoid. To obtain the ellipsoidal heights of the mountain peaks, the geoid ondulation was taken into account (ellipsoidal height = orthometric height + geoid height). For this purpose, by using an integrated function in Matlab, values between -28.98 m and -27.34 m were obtained for the Geoid height corresponding to the

Himalaya peaks. The computed RMSEs for the 7 investigated peaks register values of 10.13 m, 9.49 m, and 7.92 m in Spring 2015, Autumn 2015, and Spring 2016, respectively (Table 4.5).

The overall elevations and discrepancies between the reference ellipsoidal and reconstructed heights are shown in Figure 4.17. Generally, the absolute height differences between the satellite-driven DSMs and reference values are lower than 11 m, with the minimum and maximum values of -17.04 m and 12.54 m corresponding to the Gyachung Kang and Everest peaks in Spring 2015 and Spring 2016 datasets, respectively (Table 4.8). The world's highest mountain peak, Everest, has a reconstructed height of 8830.92 m in Spring 2015, which is higher with 12 m than the true value of 8819 m. In contrast, the ALOS model underestimates the Everest elevation to only 8768 m (51 m difference). The ALOS elevation model is being used as reference in the assessment, however big height differences occur, because of different acquisition time (2006 to 2011 ALOS mission) and lower resolution, with averaged elevations in every 30 to m pixel.

Table 4.8: Himalaya elevation discrepancies at highest peaks.

Mountain peaks (ellipsoidal heights)	Reconstructed heights (m)			Discrepancies (m)			
	Spring 2015	Autumn 2015	Spring 2016	Spring 2015	Autumn 2015	Spring 2016	ALOS 2021
Gyachung (7925 m)	7907.54	7920.56	7918.12	-17.04	-4.02	-6.46	11.42
Cho Oyu (8161 m)	8151.40	8170.95	8165.78	-9.26	10.29	5.12	-21.12
Pumori (7133 m)	7126.70	7123.44	7130.25	-6.07	-9.33	-2.52	1.55
Lobuche (6090 m)	6100.61	6102.47	6100.48	10.22	12.09	10.10	13.68
Nuptse (7826 m)	7829.93	7837.05	7834.98	3.70	10.82	8.75	-12.47
Lhotse (8487 m)	8494.18	8490.53	8492.45	7.16	3.52	5.43	-19.68
Everest (8 819 m)	8830.92	8831.34	8831.80	11.66	12.08	12.54	-51.26

The seven mountain peaks were not used as GCPs for the Himalaya dataset, because of their variable height dependent with the yearly season and also because of the difficulty in identifying their position with high accuracy in the satellite images during point measurement. Therefore they were employed only as checking points in the elevation assessment analysis.

4.2.4 Influence of acquisition geometry on DSM accuracy

To analyze the impact of the acquisition geometry on DSM accuracy, for Allentsteig (Pléiades and WorldView-3) were generated four DSMs: one using the three images simultaneously (FNB) and three using two image combinations (FN, NB and FB). The vertical quality of the resulted DEMs from each image combination were evaluated against the available elevations of the ground checkpoints and the reference lidar DTM in the open areas. Thus, the vertical accuracy in open areas was determined by (1) computing and visualizing the RMSEs between reference and obtained Z coordinates in each generated DEM and (2) analysing the histogram of difference values that should be around zero. For this, the distribution histogram was investigated by deriving statistic parameters such as mean, standard deviation, a robust standard deviation (σ_{MAD}) and also RMSE. After a quality checking, systematic errors were visible between the reconstructed DEMs and the lidar DTM. In order to reduce them, in a first attempt, the global LSM technique is applied, that estimates an affine 3D transformation between each pair of target DEM and the reference

DTM over common open areas. Hence, only 37.8% and 32.8% were used from the whole Pléiades and WorldView-3 scene extends for estimating the transformation.

Subsequently, the resulted LSM parameters were used to transform each Pléiades and WorldView-3 point clouds, which were then re-interpolated into new DEMs. Both DEM interpolation and geolocation improvement were conducted in the scientific software OPALS (Orientation and Processing of Airborne Laser Scanning Data) (Pfeifer et al. 2014).

4.2.4.1 Quality assessment of Pléiades DSMs

For each Pléiades image combination, the vertical quality of the photogrammetrically derived DEMs is assessed using the available ground check points, before and after applying the LSM transformation (Table 4.9). The RMSE values before LSM for GCPs vary between 0.95 m (FNB) and 1.02 (NB). The LSM transformation significantly reduced the RMSE values for both GCPs and CPs in each image combination, the smaller value being of 0.19 m (FNB).

Table 4.9: RMSE values in Z-direction for GCPs and CPs for Pléiades DEMs.

| No. of GCPs / CPs | RMSE (meters) | | | | | | | |
| | Before LSM | | | | After LSM | | | |
	FNB	FB	FN	NB	FNB	FB	FN	NB
43 GCPs	0.95	0.93	0.97	1.02	0.19	0.20	0.23	0.25
50 CPs	0.91	0.89	0.92	0.98	0.25	0.26	0.32	0.34

Finally, the interpolated DEMs in open, free areas for each combination, were compared with the ALS reference data. From a visual inspection, after the LSM transformation, a good agreement between DEMs and reference DTM was obtained (Figure 4.18). The statistic results show that the Dense Image Matching (DIM) terrain heights are systematically shifted by 1 m, but they are reduced to a median close to zero after applying the LSM transformation. The FNB and FB image combinations show a slightly better accuracy than the FN and NB (Table 4.10).

Table 4.10: Accuracy assessment of Pléiades DEMs in open areas (given values are in meters).

| Scene Comb. | Before LSM | | | | After LSM | | | |
	Mean	Std	σ_{MAD}	RMS	Mean	Std	σ_{MAD}	RMS
FNB	0.80	0.53	0.51	0.96	0.15	0.60	0.50	0.61
FB	0.77	0.53	0.51	0.93	0.13	0.58	0.50	0.60
FN	0.72	0.65	0.68	0.97	0.13	0.75	0.72	0.76
NB	0.78	0.70	0.73	1.04	0.17	0.79	0.80	0.81

When comparing the resulted DEMs to the ground truth lidar DTM, the elevation differences show an undulation similar to waves that are visible in the along track direction (Figure 4.19), showing an error in the photogrammetrically derived DSMs. This problem cannot come from the reference lidar DTM, because the three stereo combination differences show undulations with a wavelength of 4 500 m (Figure 4.20). In this case, LSM transformation slightly reduces the maximum difference elevation values. Except the NB-FN difference model, the wave amplitude values reach 1.5 pixels.
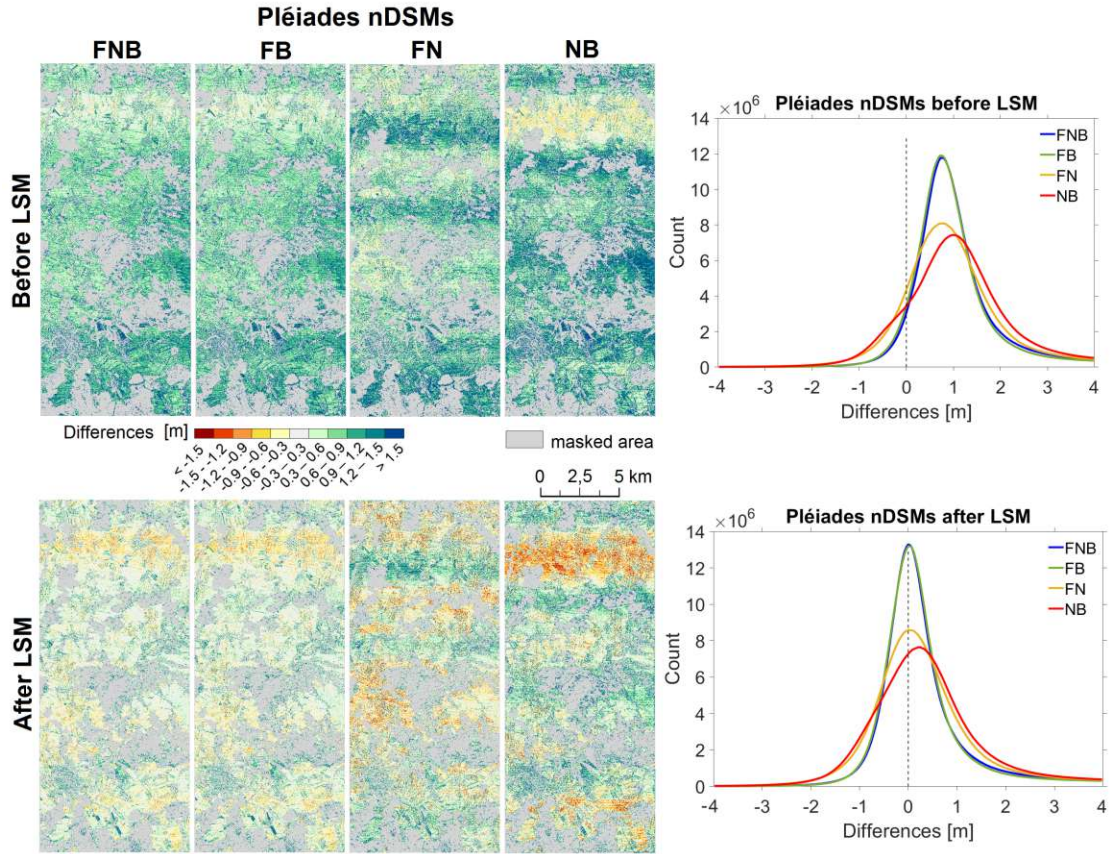
Figure 4.18: Pléiades nDSMs for open areas, before (top) and after (bottom) LSM. Left: color-coded height differences (masked areas are shown in grey). Right: frequency distribution of height discrepancies for all combinations.
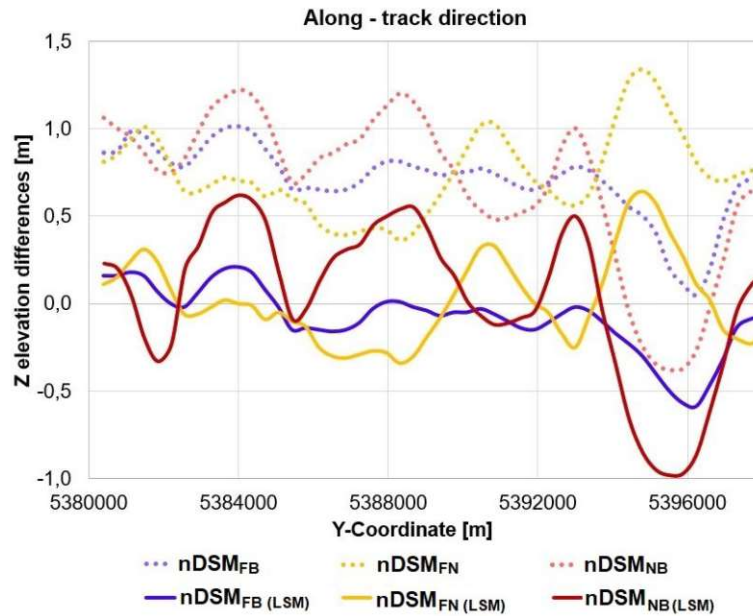

Figure 4.19: Analysis of a section profile in along-track direction with Z elevation differences before (round dot lines) and after (solid lines) LSM for the Pléiades nDSMs.
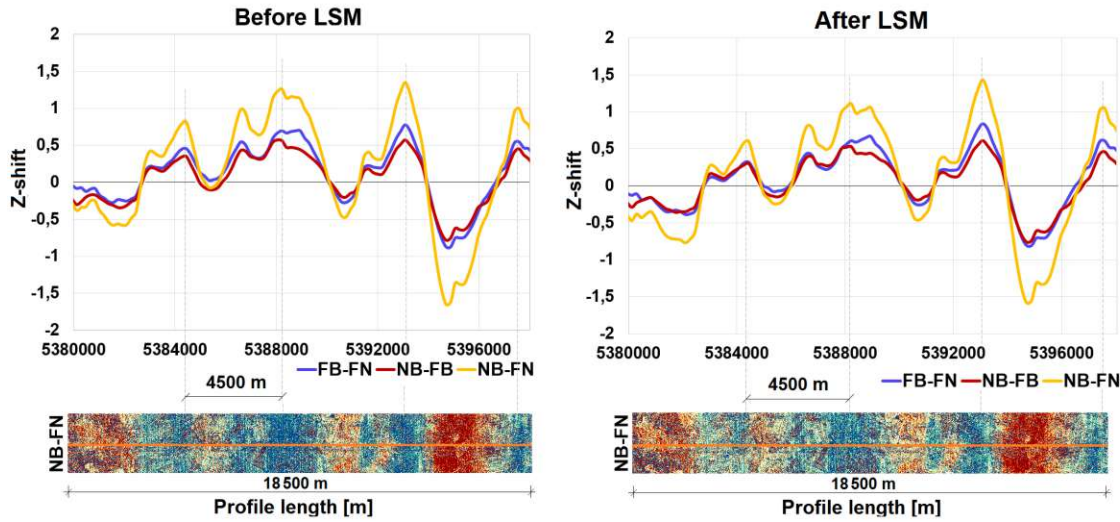
Figure 4.20: Analysis of a section profile in along-track direction before and after LSM between pairwise difference models from Pléiades satellite imagery.

4.2.4.2 Quality assessment of WorldView-3 DSMs

In a first phase, the WorldView-3 DEM accuracy is estimated by contrasting the GCPs and CPs elevations with those extracted from the computed DEM (Table 4.11), with improved results after LSM (best accuracy for GCPs in the FNB combination, of 0.12 m). The resulted statistics for CPs have higher values when compared with GCPs, due to the influence of DTM accuracy ($\sigma_z = 0.12$ m), since this served as base for CPs height extraction.

Table 4.11: RMSE values in Z-direction for GCPs and CPs for WorldView-3 DEMs.

No. of GCPs /	RMSE (meters)							
CPs	Before LSM				After LSM			
	FNB	FB	FN	NB	FNB	FB	FN	NB
36 GCPs	0.33	0.35	0.37	0.41	0.12	0.13	0.14	0.17
50 CPs	0.35	0.37	0.40	0.43	0.16	0.16	0.18	0.20

Finally, the interpolated DEMs in open, free areas were compared with the ALS reference data. Like for the Pléiades analysis, here also the pairwise comparisons allow the computation of statistic parameters. The resulting non-Gaussian distributions of the normalized elevations (nDSM) for the tri-stereo and stereo images are shown in Figure 4.21 before and after LSM. Initially with two peaks and a RMSE of 0.47 m, the histogram corresponding to NB combination is corrected by applying the LSM transformation, achieving a lower value of 0.34 m. Better and similar results are obtained for the FNB and FB image combinations after LSM (with RMSEs of 0.24 m), with lower dispersions around zero.

Similar to Pléiades, by a visual inspection of the computed nDSMs, the same wave effect in the along-track direction is visible, with a wavelength of 1500 m and a maximum amplitude of 1.5 pixels for the FB-FN and NB-FB difference models (Figure 4.22). By applying the 3D transformation, this effect is reduced to less than 1 pixel, except for the NB-FN difference model.

Another interesting effect visible in the WorldView-3 satellite-based difference ele-

Table 4.12: Accuracy assessment of WorldView-3 DEMs in open areas (given values are in meters).

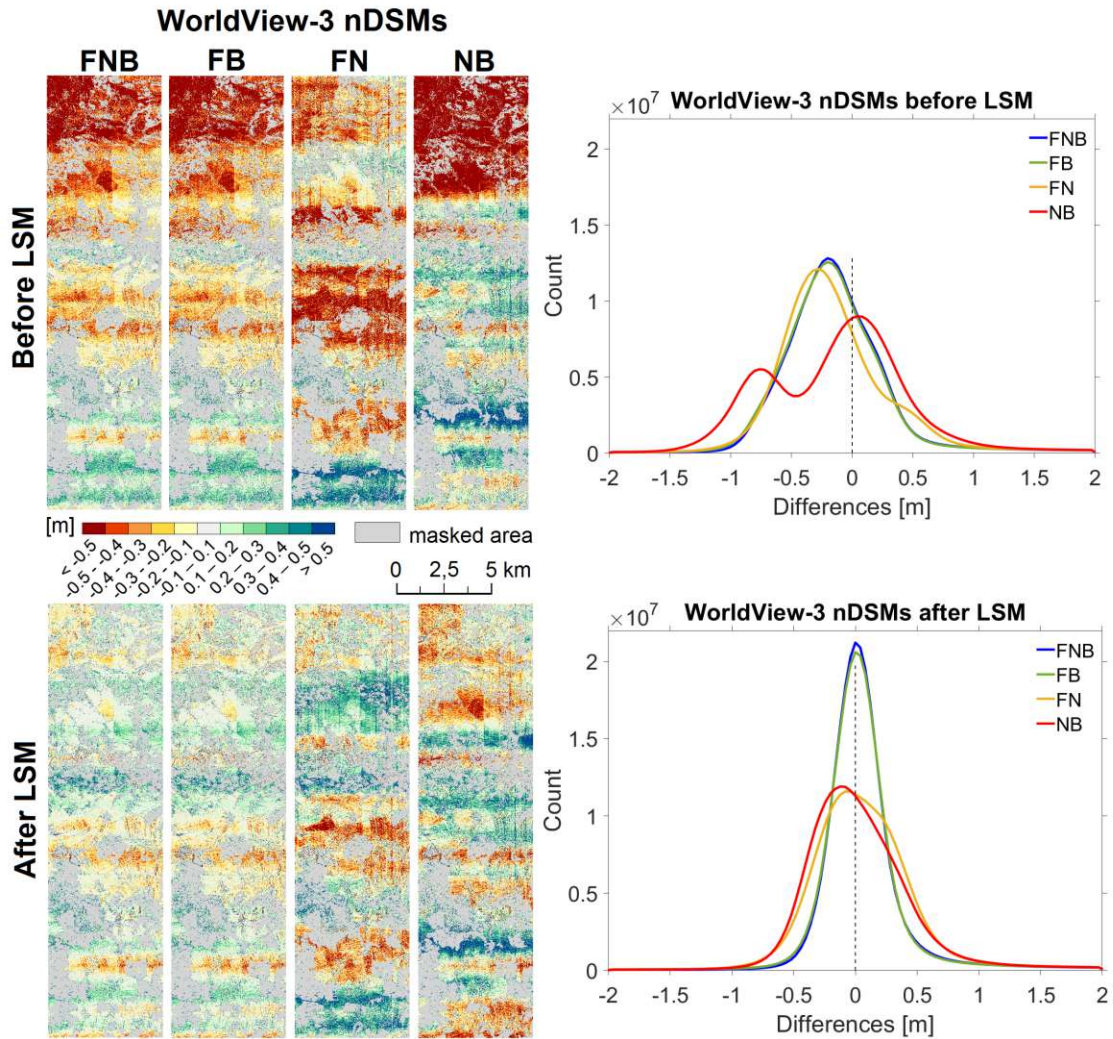| Scene Comb.	Before LSM				After LSM			
	Mean	Std	σ_{MAD}	RMS	Mean	Std	σ_{MAD}	RMS
FNB	-0.16	0.33	0.33	0.37	0.03	0.24	0.19	0.24
FB	-0.17	0.34	0.33	0.38	0.02	0.24	0.20	0.24
FN	-0.17	0.37	0.34	0.41	0.03	0.33	0.34	0.33
NB	-0.11	0.46	0.48	0.47	0.01	0.33	0.34	0.34

Figure 4.21: WorldView-3 nDSMs for open areas before (top) and after (bottom) LSM. Left: color-coded height differences (masked areas shown in grey). Right: frequency distribution of height discrepancies for all combinations.

vation models is a regular and parallel line-pattern in flying direction. This effect can be correlated with the imaging geometry of the WorldView-3 modern satellite system configuration based on CCD-line sensors (Figure 4.23).

For reaching a sufficient swath width and due to the required characteristics, a combination of shorter CCD-lines is used instead of one longer CCD-line. The satellite images collected by the individual CCD-lines do overlap slightly. The satellite cam-
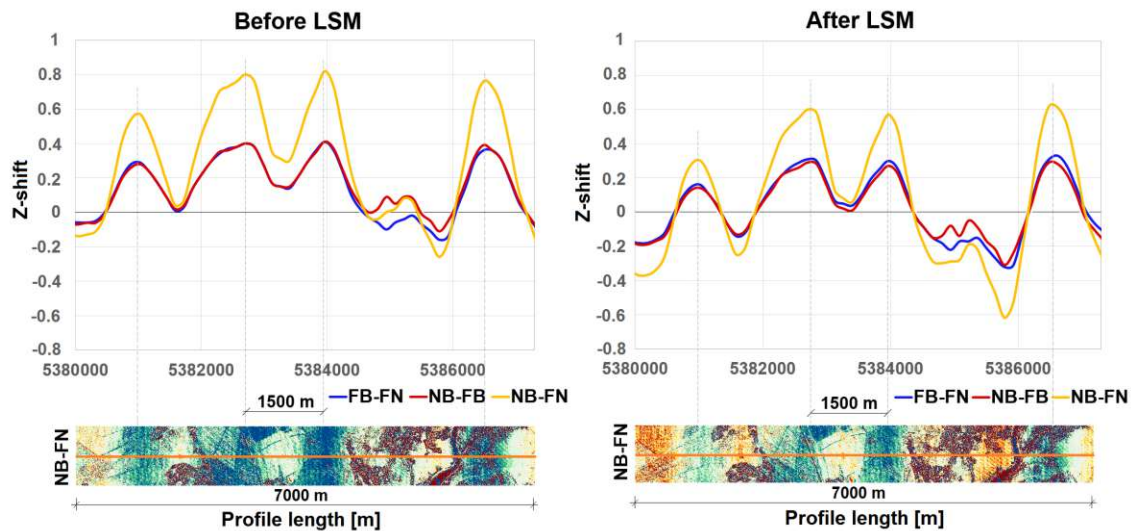
Figure 4.22: Analysis of a section profile in along-track direction before and after LSM between pairwise difference models from WorldView-3 satellite imagery.
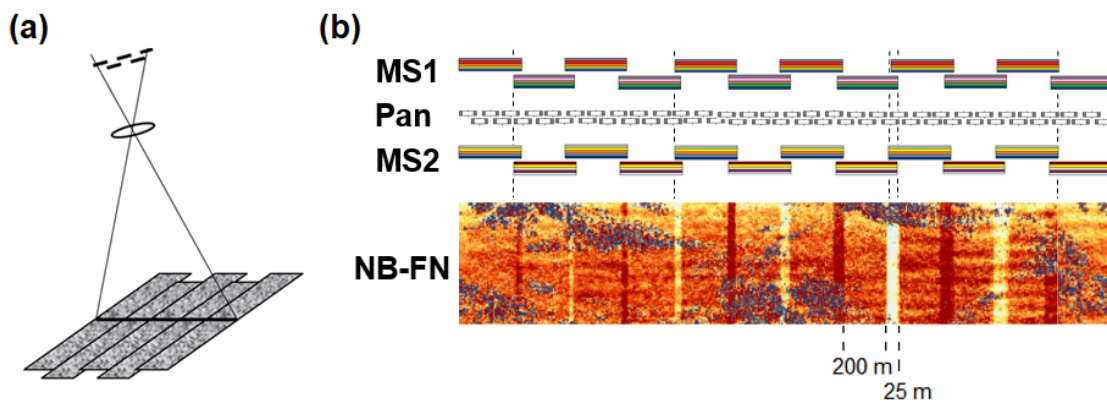

Figure 4.23: WorldView-3 sensor configuration: (a) image acquisition with staggered pushbroom array; (b) CCD configuration (Pan - panchromatic; MS1, MS2 - multispectral; adapted from Karsten Jacobsen 2005).

eras are calibrated before launch, but it cannot be guaranteed that the geometry does not change during launch or might be influenced by the thermal influence of the sun in the orbit. The exact position of CCD-lines in the camera might be slightly changed because of the large acceleration during launch. Additionally, the location of the CCD-lines for multi-spectral images needs to be determined in relation to the panchromatic CCD-line combination. Since the CCD elements cannot touch each other without a gap, they are divided into 2 groups that are offset from each other in the recording direction. In order to be able to put together a single line, the partial images from slightly different recording times must be combined with one another. The interior orientation is therefore also dependent on the exterior orientation, which can become noticeable through a slightly changed recording center and different recording directions. A periodic verification is needed, but an entire system calibration is very complex due to sensor configuration, since only the panchromatic lines of WorldView-3 are composed of 60 CCD elements. For this reason, Jacobsen

proposes a calibration by means of ground control points and overlapping scenes (Karsten Jacobsen 2005).

The existing stripes in the World View-3 elevation model of differences (NB-FN) speak of a systematic error, caused by an imperfect calibration of the CCD detectors in the array. The distances of 200 m between the lines and the overlap of 25 m on the ground correspond to 667 pixels and 83 pixels in image space, respectively.

4.2.5 Pléiades and WorldView-3 DSMs comparison

The common overlapping area in Allentsteig (of 44.5 km^2) between Pléiades and WorldView-3 acquisitions allows a comparative overview of the two sensors regarding the 3D reconstruction capability. In the overlapping area, the Pléiades-based point cloud consists of 169 million points, whereas the WorldView-3-based one is almost three times denser, with 476 million points.

A visual comparison between the DSMs from Pléiades and WorldView-3 over a rural and forested area is shown in Figure 4.24. As visible, the appearance of houses in the shaded WorldView-3 DSM have better defined geometric shapes with sharp edges. In the open fields of the Pléiades DSMs some noisy effects are visible, whereas the aspect of the forested areas is much smoother compared with WorldView-3.
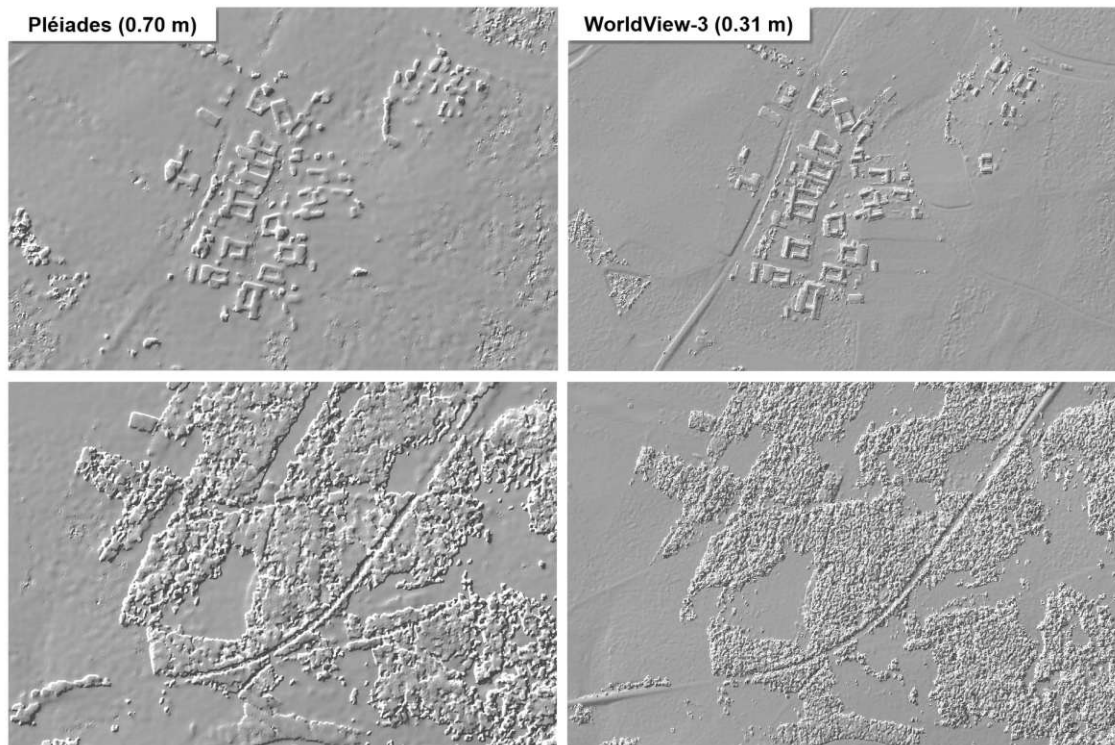

Figure 4.24: Comparison between Pléiades and WV-3 DSMs over a rural and forested area.

For obtaining statistic measures, the difference model between Pléiades and WorldView-3 DSMs was computed and a LSM transformation was applied, as well. The distribution histogram in open areas before LSM shows a positive shift, with a mean of 1.35 m, a robust standard deviation of 0.47 m and a RMSE of 1.43 m, whereas after transformation the statistic parameters decrease to 0.02 m, 0.43 and 0.93 m.
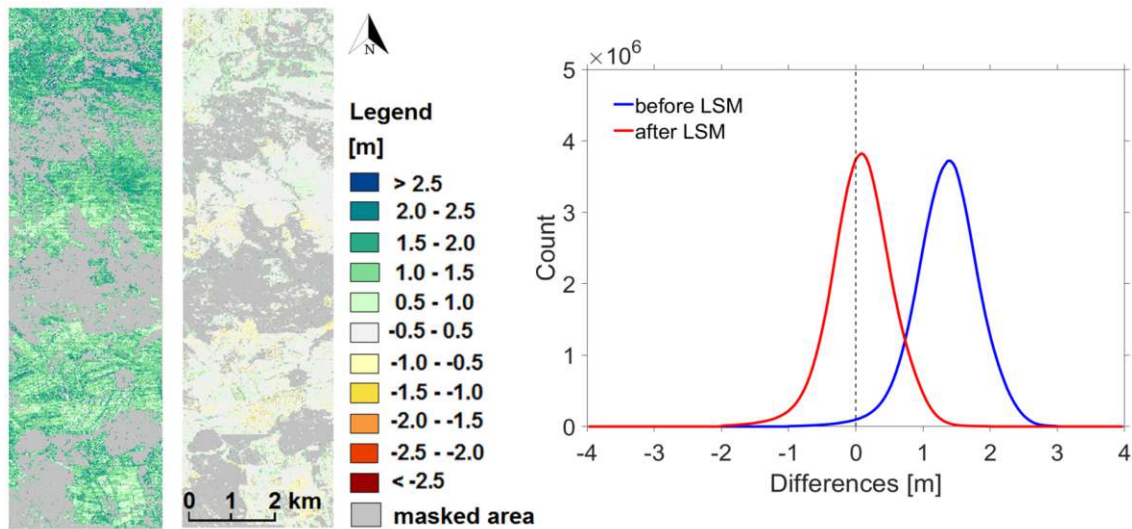
Figure 4.25: Height differences between Pléiades and WV-3 tri-stereo DEMs with distribution histograms before and after LSM (masked areas shown in grey).

When individually investigated, the quality analysis of Pléiades and WorldView-3 DEMs showed systematic errors, as well as non-normal distributions and a wavy effect in the along-track direction (sub-sections 4.2.4.1 and 4.2.4.2). Apparently the present solution of RPC refinement performed within Match-AT software is not able to fully exploit the benefit of given GCPs. Even if an additional 3D LSM transformation is applied, the wavy effect remains still present. Therefore, a further improvement is required (described in Section 4.3).

When comparing the two sensors regarding the pixel size, the application of the affine 3D LSM transformation with respect to lidar brought improvements in the vertical accuracy of the tri-stereo DEMs from 0.96 m (1.4 pixels) to 0.61 m (0.9 pixels) for Pléiades and from 0.37 m (1.2 pixels) to 0.24 m (0.7 pixels) for WorldView-3. For both sensors, the computed RMSE values vary from 0.77 to 1.15 GSD after applying the LSM transformation. These are comparable with the reported vertical accuracies for airborne photogrammetrically derived DEMs, which are between 0.44 and 2 GSD (Bühler et al. 2012, Hobi et al. 2012, Hoffmann et al. 2000, W. Hu et al. 2008, Ressl, Brockmann, et al. 2016). When compared at 50 CPs, the elevation accuracies of the Pléiades DEMs (RMSEs of 0.25 m, 0.26 m, 0.32 m and 0.34 m corresponding to FNB, FB, FN and NB, respectively (Table 4.9)) are similar with the results obtained by Bernard et al. (2012), who achieved a RMSE of 0.49 m at 295 ground control points.

Table 4.13: Vertical accuracy assessment of Pléiades and WorldView-3 tri-stereo digital surface models at the available GCPs and CPs (μ and σ are the mean and standard deviations). All values are given in meters.

Site / Sensor	No. GCPs/CPs	μ	σ	RMSE
Allentsteig	43 GCPs	0.07	0.18	0.19
Pléiades	50 CPs	0.07	0.25	0.25
Allentsteig	36 GCPs	0.03	0.11	0.12
WorldView-3	50 CPs	0.01	0.15	0.16

The accuracy values fit very well with the results reported in literature. The com-

puted mean values for the CPs of 0.07 m (Pléiades) and 0.01 m (WorldView-3) and the standard deviations of 0.25 m (Pléiades) and 0.15 m (WorldView-3) (Table 4.13) are comparable with the results obtained by Rupnik et al. 2018 who reported mean values ranging between 0.08-0.29 m and standard deviations between 0.22-0.53 m. Perko, Raggam, Gutjahr, et al. (2014) assessed the mapping potential of Pléiades stereo/tri-stereo scenes, leading to accuracies in the range of 0.5 m in planimetry and of 1 m in height. After coregistration procedures based on cross-correlation and iterative closest point, Zhou et al. (2015) reported a 0.5 m standard deviation of heights, when comparing the Pléiades DEM with an airborne lidar DEM. In accordance with the investigations reported by Piermattei, Marty, Karel, et al. (2018), higher accuracies are obtained for the tri-stereo and forward-backward combination compared with the other two stereo pairs that perform worse. The results obtained by Panagiotakis et al. (2018) over the Athens metropolitan area show that the DSM produced with tri-stereo has better performances in terms of RMSEz (1.17 m), compared with the stereo-pair DSMs that perform worse (RMSEz of 1.48 m for forward-backward and of 1.64 m for forward-nadir).

Compared with Pléiades, the WorldView-3 DSM has a higher vertical accuracy. This is because the vertical accuracy of the DSM is directly influenced by the triplet acquisition geometry, especially by the intersection angle on the ground. As described in Daniela Poli and Caravaggi (2013), the vertical accuracy of the DSMs from VHR satellite imagery is influenced by the acquisition geometry, where a wider angle of convergence ($> 15^o$) enhances the height accuracy. In this case, the Pléiades scenes with a narrow convergence angle (12°) show a lower vertical performance than the WorldView-3 scenes, with a larger convergence angle on the ground (23°). Moreover, the smaller ground sampling distance (0.31 m) of the WorldView-3 images leads to a more reliable representation of details on ground surface. The vertical accuracy results agree with those reported by F. Hu et al. (2016), of less than 0.5 m for 6001 ground lidar CPs. The small acquisition convergence angles influence the vertical accuracy, as the images are very similar to each other, a fact that enables a better image matching, despite the decrease in geometric intersection quality. By careful georeferencing including LSM over stable areas, change detection between Pléiades and WorldView-3 satellite images is limited by an accuracy of only 0.50 m. With accuracies comparable to airborne photogrammetrically derived DEMs, the VHR satellite images are an important alternative data source for high resolution digital elevation models derivation over large areas.

The two high resolution DSMs derived from Pléiades and WorldView-3 tri-stereo scenes within the WGS84-UTM33 reference system were further used as source data for the single object investigations.

4.2.6 DTM derivation in urban areas

For deriving the DTM from the photogrammetric DSM over Barcelona city, the volume-based filtering approach developed by Piltz et al. (2016) and implemented in the Opals software was used. The detailed workflow for DTM derivation is depicted in Figure 4.26. The mask for the open terrain was computed by setting a minimum height of the off-terrain objects to 3 m and maximum width to 80 m. The width was set in accordance with the maximum length/width of existing buildings in the Barcelona dataset. In the next step, morphological operations are applied to the
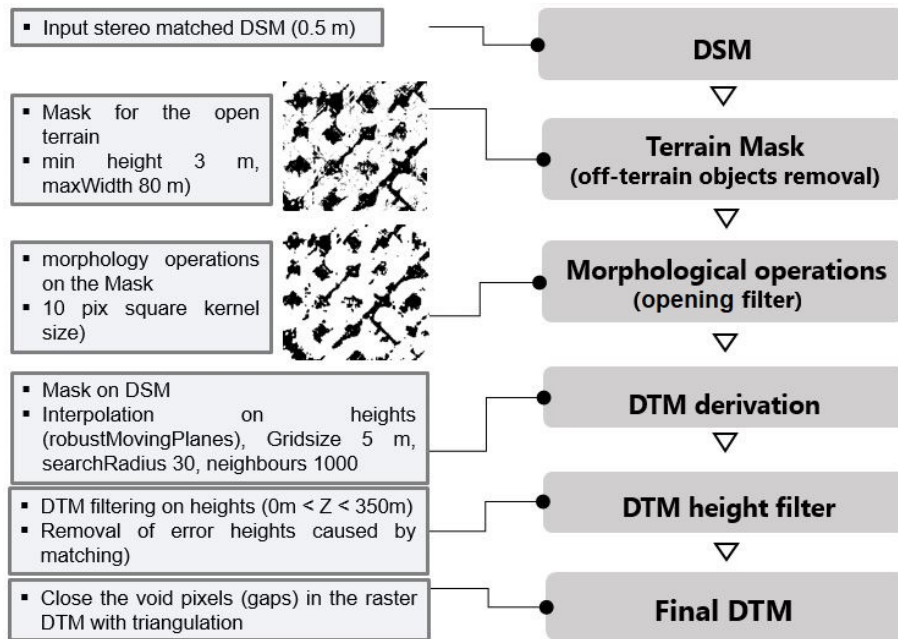
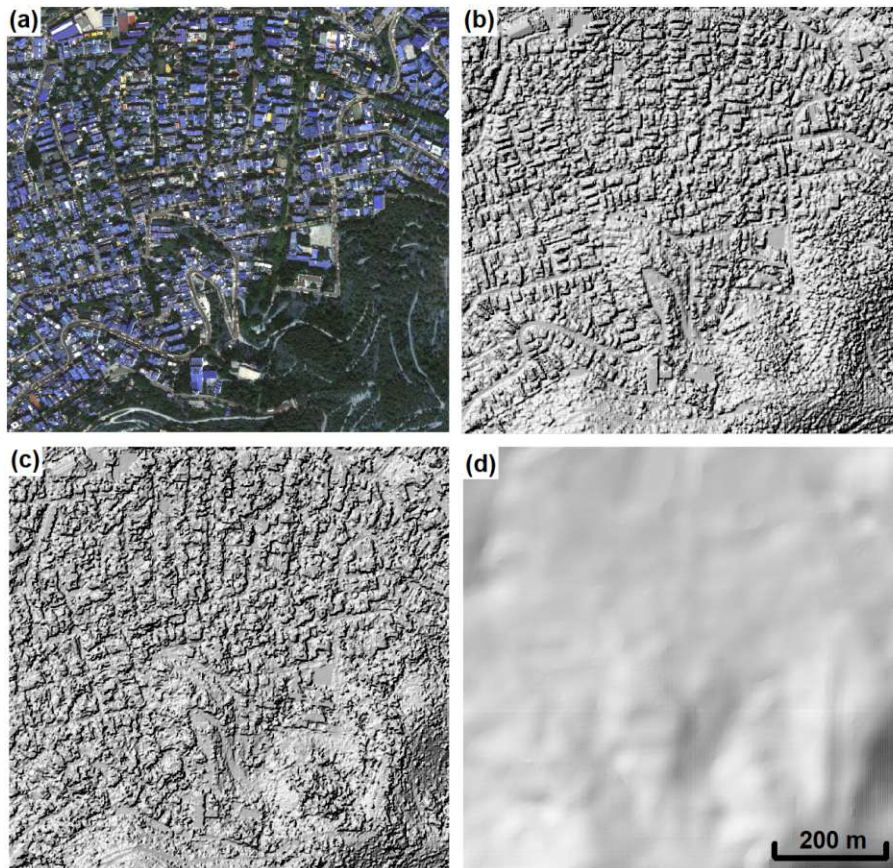Figure 4.26: Processing workflow for DTM derivation.

Figure 4.27: Zoomed-in area over Barcelona: (a) orthophoto; (b) shaded lidar DSM; (c) shaded satellite-based DSM; (d) shaded DTM.

mask. This was further used for obtaining the terrain elevations only. Finally, the areas with missing elevation information were filled by triangulation.

Figure 4.27 shows a zoomed-in area in Barcelona with visualisations of the orthofoto, shaded lidar DSM and satellite-based DSM and DTM. In comparison with lidar, the computed DSM shows a high rougness. The computed DTM has a very smooth appearance (looking even blurred), because of the high number of neighbours (1000) and large search radius (30 m) used for the interpolation.

4.3 Improvement of satellite DSMs

A detailed quality analysis of the satellite-based DSMs together with the results from the adopted improvement methods are thoroughly described in this section. Special attention was dedicated to Pléiades tri-stereo dataset over Allentsteig test area. In comparison to a stereo pair, a triple stereo scene offers the possibility to build three image pairs to compute a surface model. With perfectly calibrated and oriented images all three surface models deviate from a reference model, but also from each other only by random errors. As Figure 4.28 shows, this is not the case for the investigated data. A surface model computed from all three images shows differences to the Lidar reference model as systematic deviations with an amplitude reaching 0.88 m. These differences can only be determined, if a reference model of higher (bias free) accuracy is available. However, differences between the surface models computed between the different pairs also show systematic patterns, as demonstrated in Figure 4.29. In this case, the elevation differences for $DSM_{FB} - DSM_{FN}$, $DSM_{NB} - DSM_{FB}$ and $DSM_{NB} - DSM_{FN}$ reach amplitudes of 0.87 m, 0.78 m and 1.65 m, respectively. In following, these effects are investigated and possible remedies for improvement are proposed.
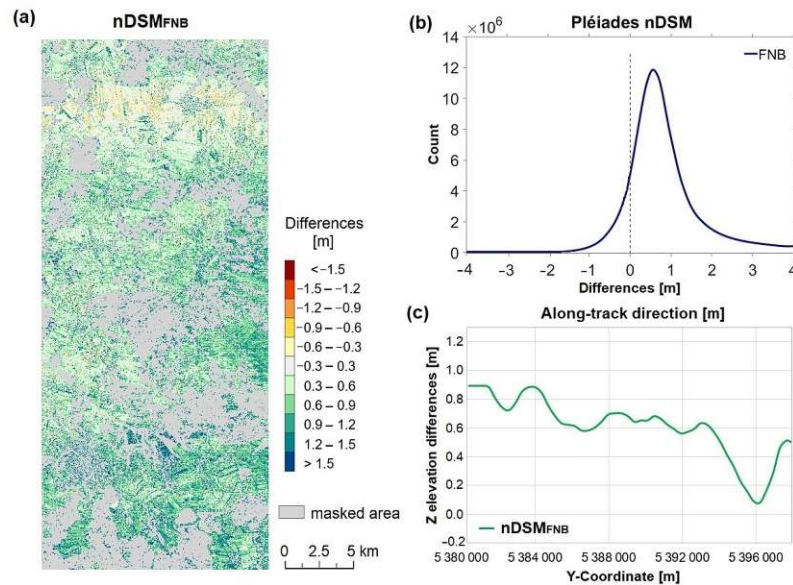

Figure 4.28: Differences between Pléiades DSM_{FNB} and Lidar reference DTM in open areas: (a) Color-coded height differences (masked areas are shown in grey); (b) Frequency distribution of height discrepancies; (c) Section profile in along-track direction with Z elevation differences.
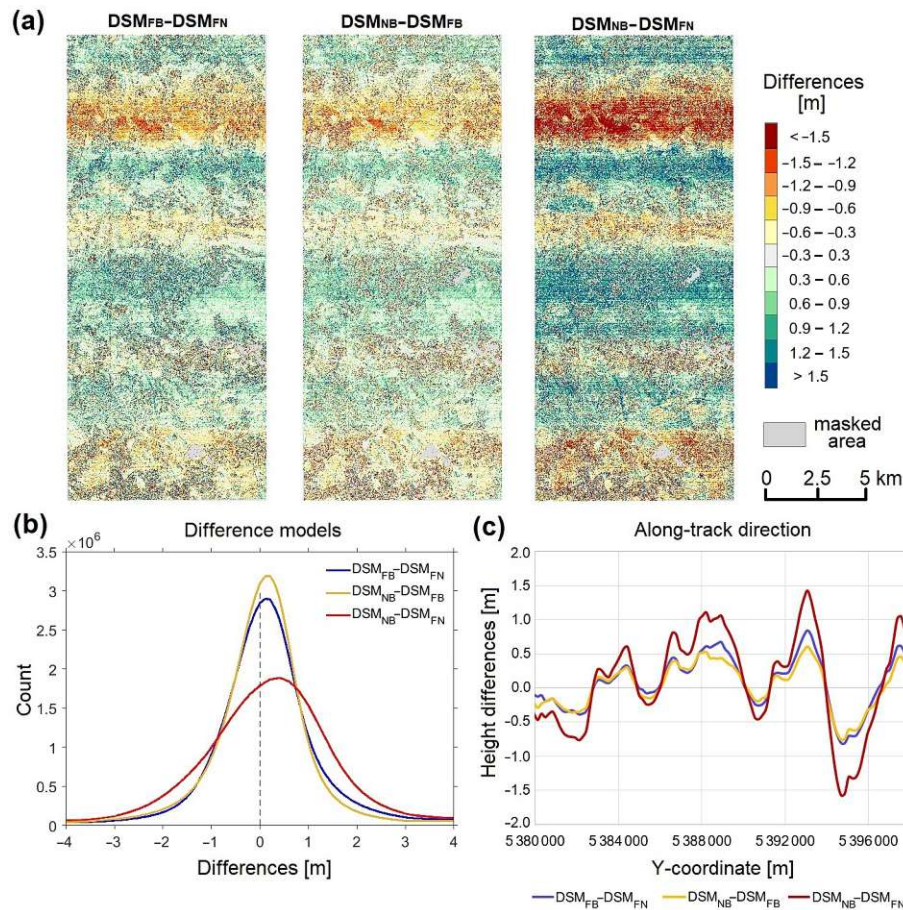
Figure 4.29: Differences between Pléiades pairwise DSMs: (a) Color-coded height differences; (b) Frequency distribution of height discrepancies; (c) Section profile in along-track direction with Z elevation differences.

4.3.1   DSM from image orientation with bias-corrected RPCs using 43 and 300 GCPs

For an accurate processing, the satellite image orientations provided through the RPCs were improved by estimating an affine transformation in image space, making use of available GCPs. Two situations were analysed:

1. Image georeferencing with bias-corrected RPCs using 43 GCPs and automatically extracted TPs.

2. Image georeferencing with bias-corrected RPCs using 300 GCPs (no TPs are involved).

The error vector plots in image space (for GCPs and TPs) resulted from satellite triangulations in the two cases are shown in Figures 4.30 and 4.31 and statistics are given in Table 4.14.

In both image block-orientations, values in the range of $1/3^{rd}$ of a pixel were achieved for the standard deviations of point residuals. When looking at the statistics for image residuals of GCPs and automatic TPs, generally the values in cross-track direction (x-residual) are larger than those in-track direction (y-residual). Assuming that the epipolar lines are approximately parallel to the flight direction, this indicates a good estimation in cross-track (cross epipolar line) direction, but less control in

Table 4.14: Pléiades tri-stereo satellite imagery aerial triangulation results.

Image	No. GCPs	No. TPs / Image	Sigma (px)	Image residual statistics for GCPs & TP obs.					
				x-residual (px)			y-residual (px)		
				μ	σ_{MAD}	RMSE	μ	σ_{MAD}	RMSE
Forward		373		0.00	0.26	0.26	0.00	0.13	0.13
Nadir	43	378	0.28	0.00	0.28	0.27	0.00	0.21	0.22
Backward		375		0.00	0.26	0.26	0.00	0.12	0.12
Forward		-		0.00	0.26	0.26	0.00	0.24	0.24
Nadir	300	-	0.32	0.00	0.40	0.40	0.00	0.38	0.38
Backward		-		0.00	0.32	0.31	0.00	0.32	0.31

flying direction, keeping some systematic errors. Moreover, when comparing the RMSE y-residual within Allentsteig image triplet - 0.13 pixels, 0.22 pixels, 0.12 pixels in the first case (RPC bias-correction with 43 GCPs) and 0.24 pixels, 0.38 pixels, 0.31 pixels in the second case (RPC bias-correction with 300 GCPs) for Forward, Nadir, Backward images - larger value for the Nadir image suggests a better estimation of the orientation in comparison to Forward and Backward scenes. This effect can be seen in parallel with the small correction values for the Nadir image in Section 4.3.3.1. No systematic component in y-direction can be observed for the image point residuals obtained based on bias-corrected RPC orientation using 43 GCPs (Figure 4.30).
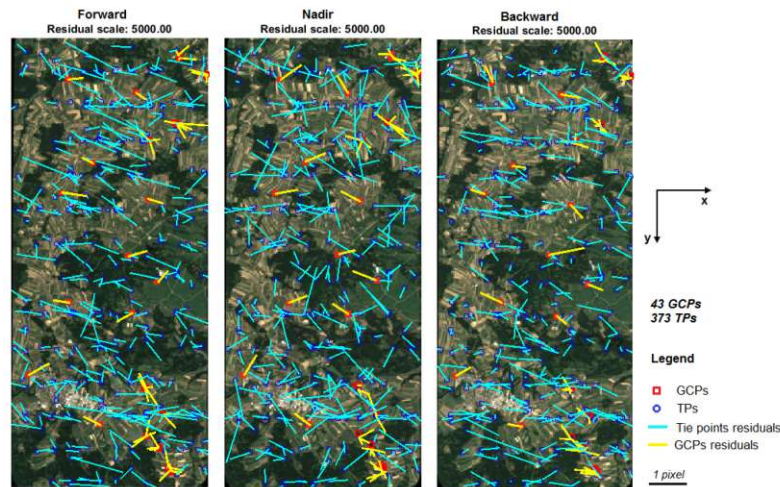

Figure 4.30: Image re-projection errors of the 43 GCPs and 373 TPs in Allentsteig dataset based on bias-corrected RPC orientation. Overview of scaled image residuals in Forward, Nadir and Backward images. The vectors were exaggerated 5000 times to enable better illustration.

In the second case, satellite image orientation was performed with bias-corrected RPCs using 300 GCPs (no TPs included). Figure 4.31 provides an insight into the question of modelling the systematic error in the along-track direction (y), by making use of residuals robust statistics, such as median and standard deviations. The basic idea was to group the y-residuals in equal intervals along track direction and check if the median values of the frequency distribution in each interval shows similar patterns, thus deviating from a purely random distribution. However the

resulted graphs per each image displaying the medians in y-direction and robust standard deviations for 8 intervals (Figure 4.31 (b)) do not show regular oscillations. While the Forward image has two complete oscillations, the Nadir graph is mostly flat and the Backward irregular. By increasing the number of intervals along track direction, the robust standard deviations remain mostly constant: between 0.1 - 0.15 pixels for Forward, smaller than 0.10 pixels for Nadir, and between 0.15 - 0.25 pixels for Backward. The high variability of medians-Y values leads to the conclusion that the 300 GCPs are not suited for identifying systematic errors from image residuals, therefore a different source, e.g. a digital elevation model is needed. More information from the satellite gyros could also help in this case. Apparently, the quite random distribution of the GCPs coordinate residuals in image space (showed as vector plots in Figure 4.31) suggests the absence of any further systematic error.
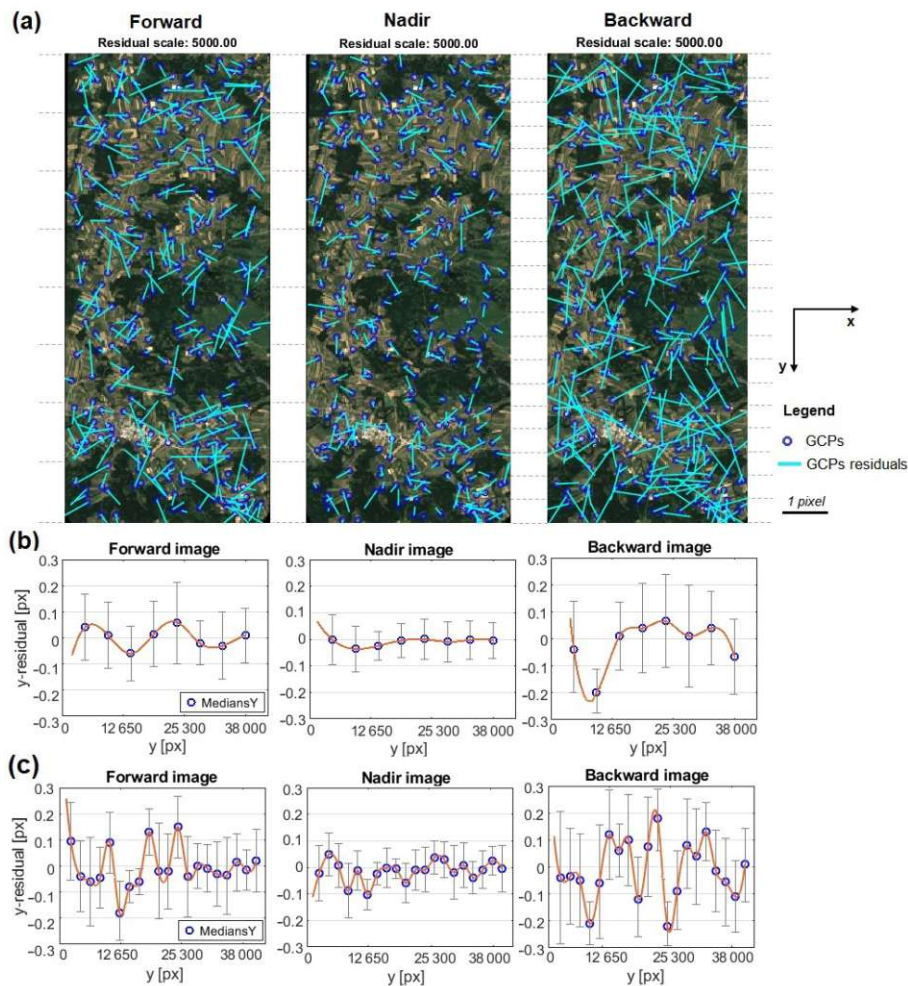

Figure 4.31: Image re-projection errors of the 300 GCPs in Allentsteig dataset based on bias-corrected RPC orientation: (a) Overview of scaled image residuals in Forward, Nadir and Backward images with marked 8 and 20 section divisions in flight direction. The vectors were exaggerated 5000 times to enable better illustration; (b) Median with corresponding robust standard deviations for 8 along-track intervals; (c) Median with corresponding robust standard deviations for 20 along-track intervals.

The accuracy of the computed satellite imagery aerial triangulation was evaluated by comparing the GCPs coordinates - measured directly on the oriented satellite images - with those measured on the terrain (or extracted from the orthophoto and lidar, after case). Statistic results for the 43 GCPs are shown in Table 4.15. Figure 4.32 (identical to Figure 4.32 top left) shows the graphs of comparisons at GCPs, where the coordinate differences are less than 1 m for most of the points.

Table 4.15: Accuracy statistics at 43 GCPs coordinates and on elevation for open areas (*Vertical accuracy of Pléiades DSM against lidar DTM in open, object-free areas).

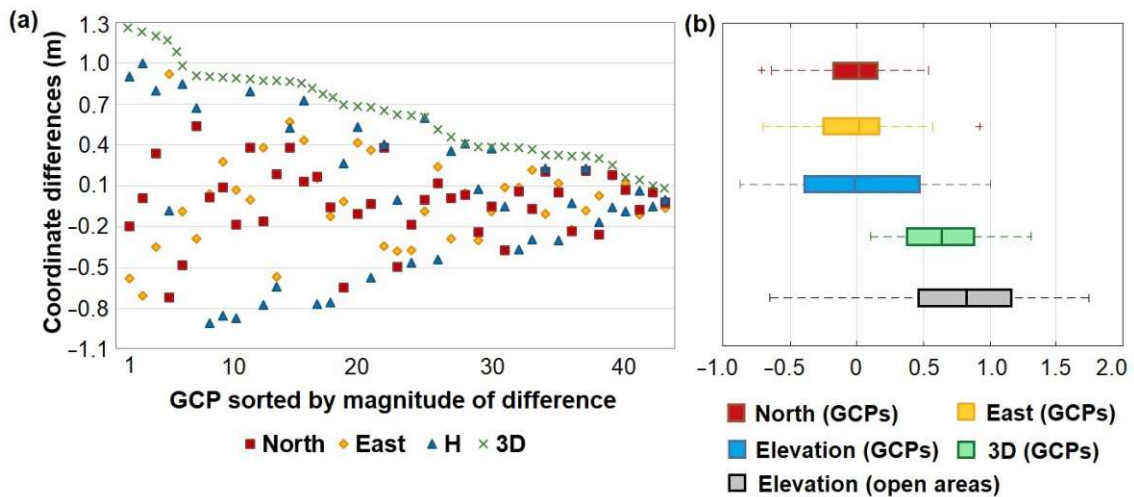| Study site	Allentsteig			
	Mean (m)	Median (m)	σ (m)	RMSE (m)
Easting	-0.01	0.02	0.29	0.32
Northing	-0.02	0.01	0.22	0.27
Elevation	0.05	-0.01	0.12	0.57
3D	0.64	0.64	0.16	0.71
***Open areas (Elevation)**	0.80	0.79	0.53	0.89

Figure 4.32: Comparisons between GCPs coordinates and monoscopically measured on the oriented satellite images: (a) Coordinate differences; (b) Statistics on GCPs coordinates and on elevations in open areas.

At a wider scale, the accuracy of the satellite-based DSMs was evaluated by comparing them with the ALS DTM reference data in open, free areas. Figure 4.33 shows a comparative view of the resulted nDSMs from the two investigations: RPCs refined with 43 GCPs and with 300 GCPs. In both cases, the color-coded height difference models reveal periodic systematic errors (smilar to an undulation effect with waves) that are visible in the along track direction. Statistics of elevation differences show non-Gaussian distributions of errors with a positive vertical offset of the Pléiades DSM (with approximately 1 m) over the reference Lidar DTM. In contrast, the elevation models computed with 300 GCPs show a better agreement with the Lidar DTM, with mean values close to zero and the best RMS of 0.65 m for the DSM from FB image combination (Table 4.16).

Table 4.16: Accuracy assessment of Pléiades DSMs in open areas (given values are in meters).

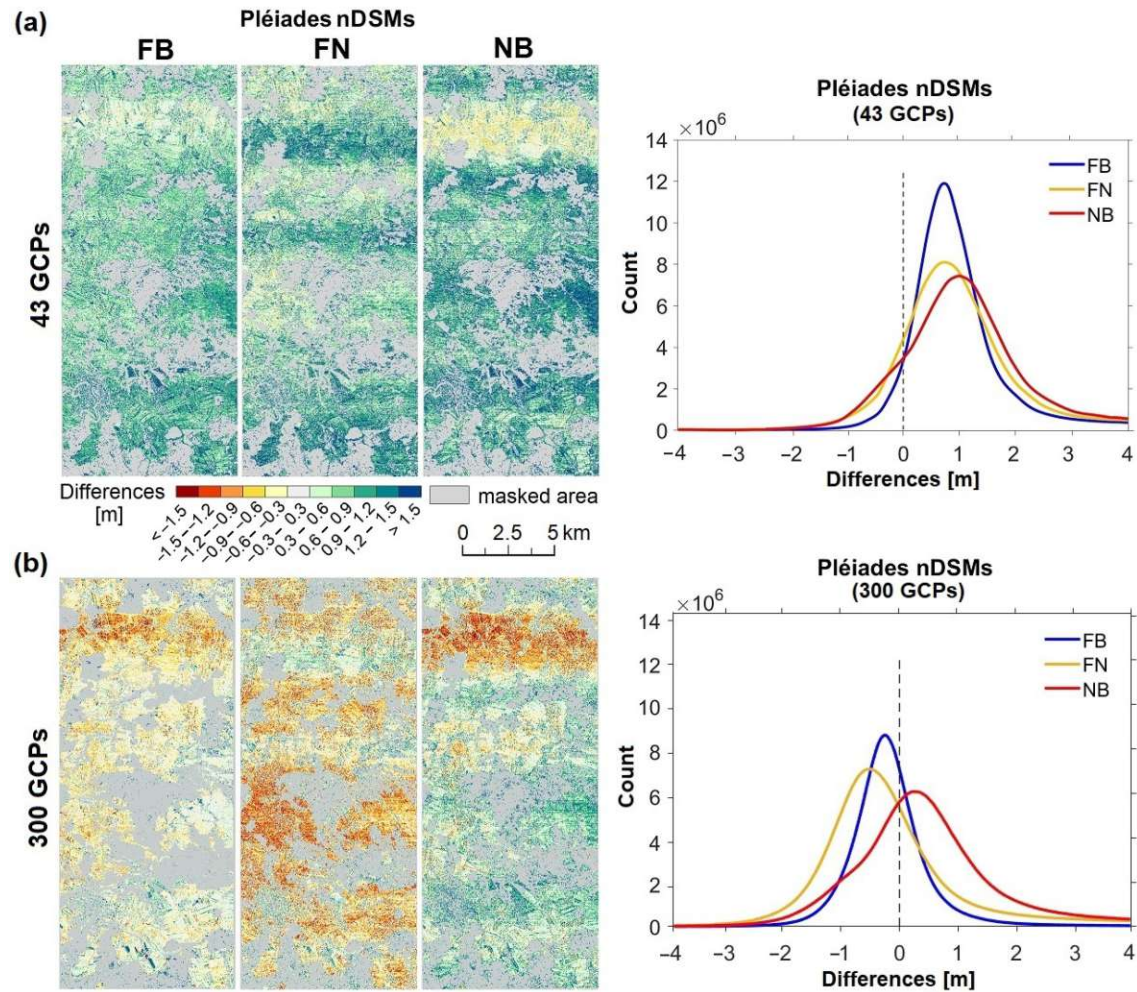| Scene Comb.	RPC refinement with 43 GCPs				RPC refinement with 300 GCPs			
	Mean	Std	σ_{MAD}	RMSE	Mean	Std	σ_{MAD}	RMSE
FB	0.77	0.53	0.51	0.93	-0.10	0.64	0.53	0.65
FN	0.72	0.65	0.68	0.97	-0.32	0.78	0.73	0.84
NB	0.78	0.70	0.73	1.04	0.21	0.83	0.82	0.85

Figure 4.33: Pléiades nDSMs for open areas with corresponding frequency distribution of height discrepancies: (a) Image georeferencing with bias-corrected RPCs using 43 GCPs; (b) Image georeferencing with bias-corrected RPCs using 300 GCPs.

4.3.2 Satellite DSMs corrected with LSM

In order to reduce the systematic errors in the satellite-based DSMs, a global LSM technique between the satellite-based DSMs and reference Lidar DTM in open areas was applied. From a visual inspection, an improved agreement between DSMs and reference DTM is obtained, although the systematic wavy effect is still preserved (Figure 4.34).

The statistics of elevation differences between the satellite-driven DSMs (all combi-
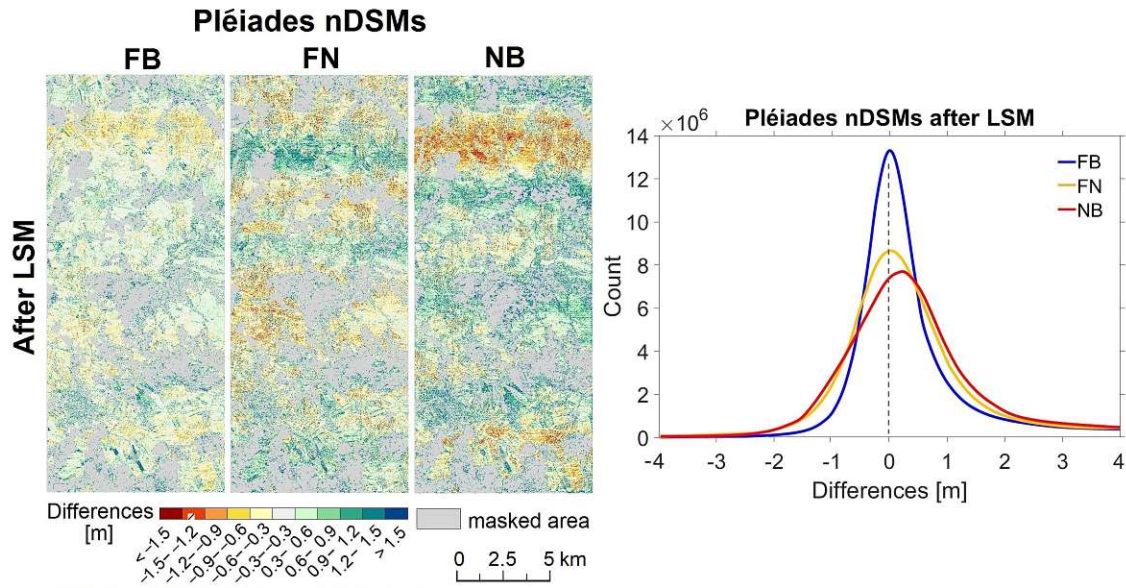
Figure 4.34: Pléiades nDSMs (RPC refinement with 43 GCPs) for open areas after global LSM. Color-coded height differences (masked areas are shown in grey) with corresponding frequency distribution of height discrepancies for all combinations (to be compared with Pléiades nDSMs before global LSM in Figure 4.33(a)).

nations) and available reference model before and after applying global LSM technique are given in Table 4.17. Initial systematically shifted by 1 m, the reconstructed terrain heights are reduced after applying the transformation. The global LSM technique substantially reduced the RMS values with 22% for FN and NB scenes, and with up to 35% for FB scene.

Table 4.17: Accuracy assessment (RPC refinement with 43 GCPs) of Pléiades DSMs in open areas (given values are in meters).

Scene Comb.	Before LSM				After LSM			
	Mean	Std	σ_{MAD}	RMSE	Mean	Std	σ_{MAD}	RMSE
FB	0.77	0.53	0.51	0.93	0.13	0.60	0.50	0.60
FN	0.72	0.65	0.68	0.97	0.13	0.75	0.72	0.76
NB	0.78	0.70	0.73	1.04	0.17	0.79	0.80	0.81

Computed with LSM continuous approach, the elevation corrections describe the offsets in Z direction as a function of Y coordinate (in flight direction) (Figure 4.35). The elevation differences from the FB surface model are around 0 m (they are the closest to 0 m, compared to NB and FN). While the maximum absolute difference for the FB model with respect to Lidar DTM is of 68 cm, the differences for NB and FN elevation models reach minimum values of -88 and of -85 cm, respectively. Similar to oscillations, wavy patterns are visible in all the three difference models, more pronounced in FN and NB, which have a symmetric appearance to the horizontal 0 m axis. The FB model is similar to NB, but with waves of smaller amplitudes.

As described in Section 2.3.3, the elevation corrections are then applied to the defined 2D positions (latitude, longitude) of 5 583 points distributed at 10 m distance interval in three parallel lines (1861 points/line).
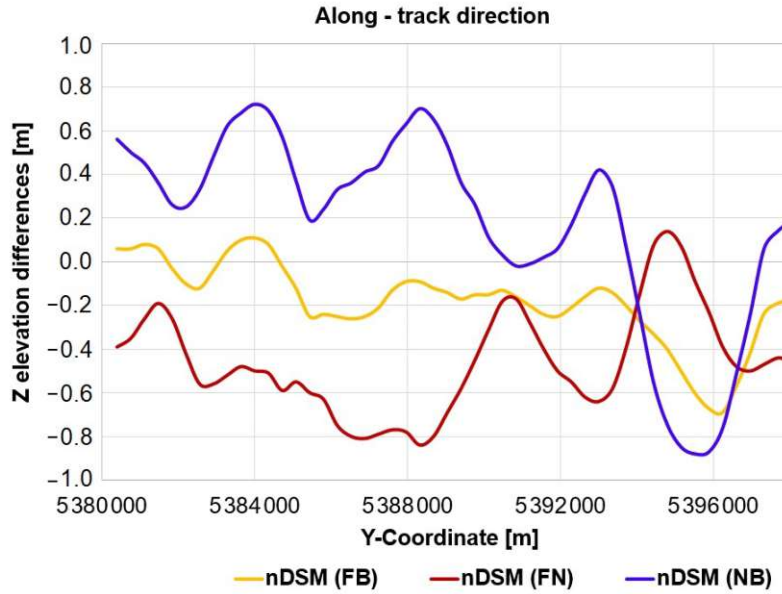
Figure 4.35: Elevation corrections shown as profile-height differences for pairwise Pléiades nDSMs.

4.3.3 Satellite DSMs corrected based on image warping

4.3.3.1 Image Correction Models

Since each satellite image is involved in the computation of two DSMs (Forward image in DSM_{FB}, DSM_{FN}, Nadir image in DSM_{FN}, DSM_{NB}, and Backward image in DSM_{NB}, DSM_{FB}), two different corrections per image are obtained. The resulted correction functions and their average for the tri-stereo Pléiades satellite images are shown in Figure 4.36. For the Forward image, the correction functions computed from the two DSMs are very similar and fit very well together, whereas for the Nadir image the amplitudes of the waves are the smallest (<0.02 pixels). The largest discrepancies are visible in the corrections corresponding to the Backward image. Even if the oscillation pattern is similar for both FB and NB DSMs, the correction function corresponding to the NB DSM has the maximum amplitude and height, with values of 0.125 and 0.23 pixels, respectively. The resulted average image corrections for the tri-stereo Pléiades satellite imagery have amplitudes of maximum 0.063 pixels, 0.020 pixels, and 0.075 pixels for the Forward, Nadir, and Backward images, respectively. Overall, the average correction functions in all images are characterized by repetitive waves with lengths of 9 509 pixels (wavelength = image rows / no. of waves, 38 034 pixels / 4 waves), amplitudes varying between 0.012 and 0.075 pixels and heights between 0.025 and 0.15 pixels (Figure 4.36(d)).

A maximum wave height of 0.15 pixels corresponds to 10.5 cm on the ground. The image correction functions can be interpreted as describing the vibrations (jitter) of the satellite during image acquisition, by moving back and forth from its normal stationary position on the platform.

The distribution histograms for the corrections show higher values for mean and standard deviation in the Backward image when compared to the other two. The corrections for Nadir image are the smallest, between -0.05 and 0.05 pixels (Figure 4.37).
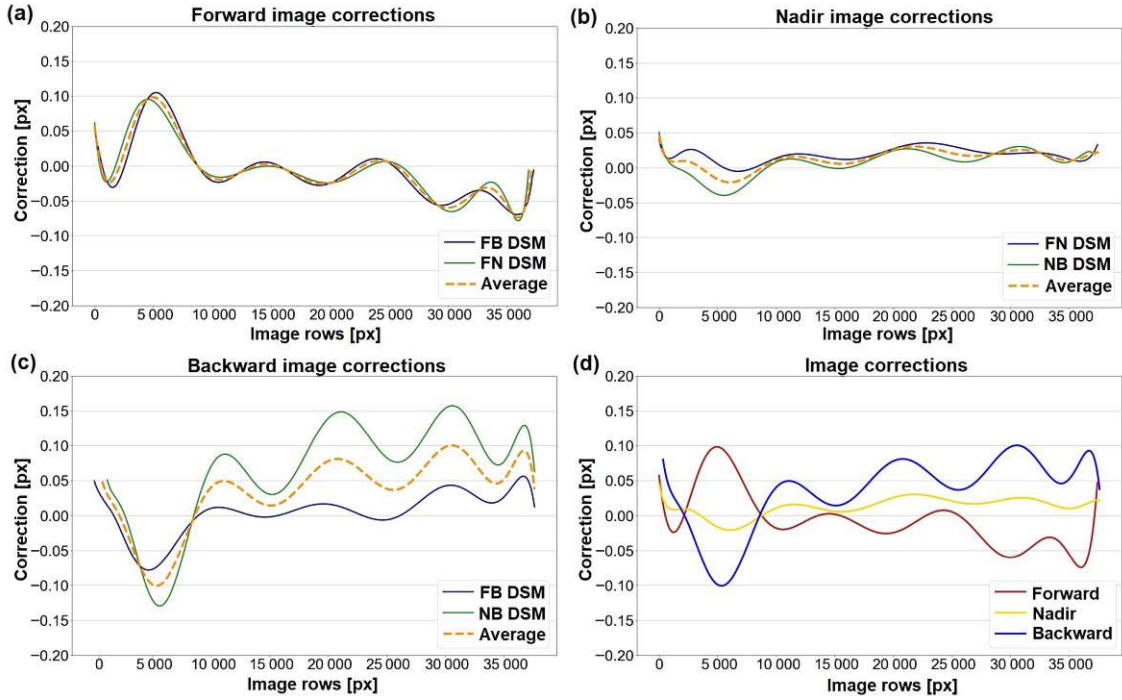
Figure 4.36: Correction functions for Pléiades satellite imagery: (a) Forward image; (b) Nadir image; (c) Backward image; (d) Averaged values.
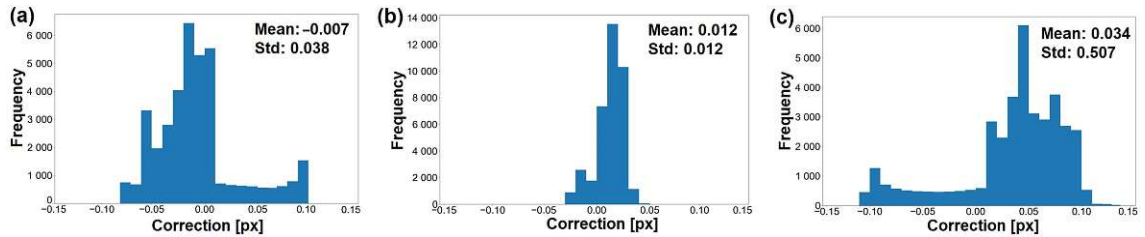

Figure 4.37: Histograms of correction values for Pléiades satellite imagery: (a) Forward image; (b) Nadir image; (c) Backward images.

When reversing the image rows (considering the opposite direction of rows from top to bottom in image coordinate system with origin in top left corner), a similarity between image correction functions (Figure 4.36(d)) and the profile height differences for pairwise Pléiades nDSMs in object space (Figure 4.35) can be observed. The four dominant peaks in the correction models for each image correspond to the four peaks in the computed normalized difference surface models.

4.3.3.2 Sensor oscillations

Describing oscillations with wavelengths of 9 509 pixels, the determined image corrections reveal important technical information regarding the image acquisition w.r.t sensor rotation in time (pitch angle) and satellite oscillations. The sensor vibration i.e. change of the viewing direction from the sensor towards ground for each wavelength can be determined by the following formula:

$$\vartheta = \arctan \frac{2 \cdot y}{H} \tag{4.1}$$

where y is considered the maximum wave amplitude of 0.15 px (0.105 m) and H the Pléiades satellite orbit height of 694 km. This results in a very small oscillation angle of the sensor with a value of $1.7 \cdot 10^{-5}$ degrees.

Based on the satellite image acquisition properties also the time for a complete cycle of the oscillation (oscillating period T) can be computed. One satellite image (16 665 x 38 034 pixels) is acquired in less than 3 seconds (2.795 s Forward, 2.823 s Nadir, and 2.803 s Backward images), which typically results in a recording time of $7.36 \cdot 10^{-5}$ for each row-line. An oscillation with 9 509 pixel wavelength (corresponding to 6 656 m on the ground) is recorded in a time interval of 0.70 s. Knowing that frequency f refers to the number of waves per unit of time, given by f = 1/T, the satellite vibrations have a frequency of 1.42 Hz.

4.3.3.3 Evaluation of DSMs after image geometry correction

The distortions in the satellite images were corrected by using the warping technique and together with the GCPs were finally used as input in a new photogrammetric processing chain for deriving elevation models.
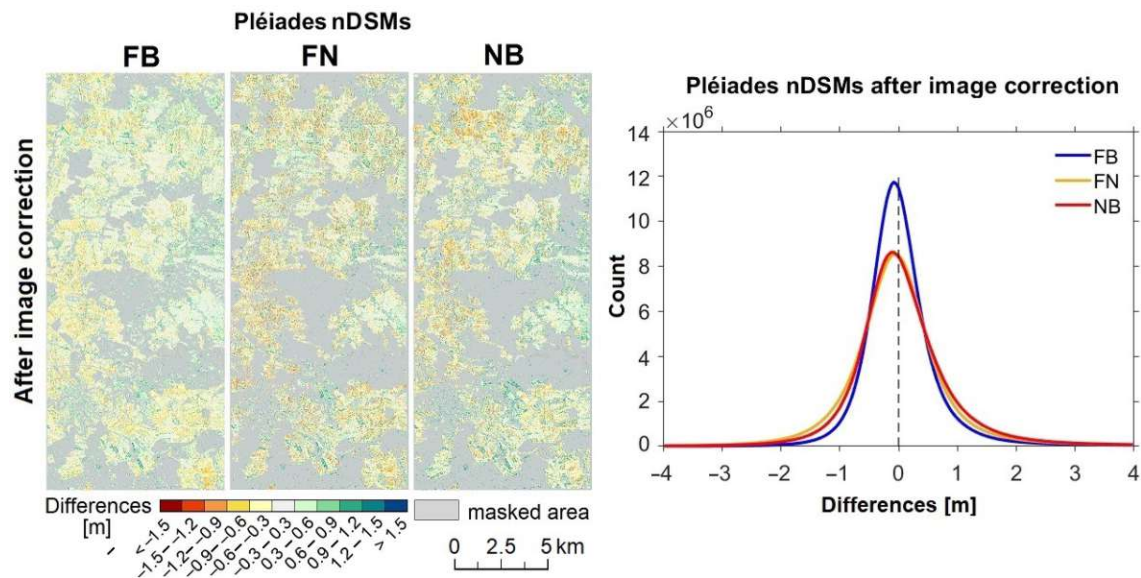

Figure 4.38: Pléiades nDSMs for open areas after image correction. Color-coded height differences (masked areas are shown in grey) with corresponding frequency distribution of height discrepancies of Pléiades DSMs against reference Lidar DTM (to be compared with Pléiades nDSMs before image correction in Figure 4.33).

The new interpolated satellite-based DEMs in open, smooth areas for each stereo pair were again compared with the ALS reference model. As visible in Figure 4.38, the systematic elevation differences (with oscillation pattern) are successfully removed by applying the image correction functions. The section profile in along-track direction with Z-elevation differences of pairwise DSMs against reference Lidar DTM shows values close to 0 m after correction (Figure 4.39). Initially with maximum elevation differences of up to 0.88 m (Figure 4.35), the improved DSMs show less than 0.10 cm differences after correction.

Table 4.18 shows the vertical differences statistics of Pléiades-based DSMs before and after applying image distortion correction. Initially systematically shifted (mean
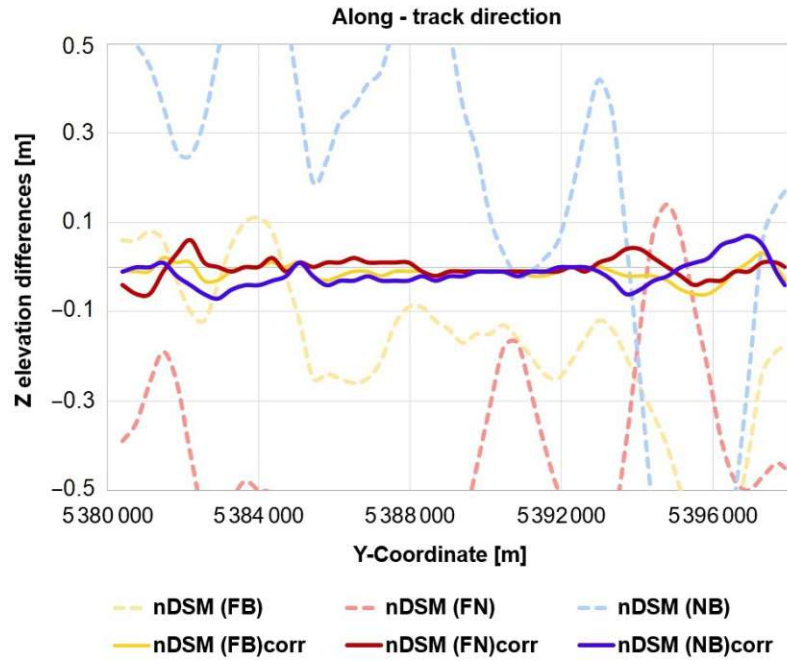
Figure 4.39: Section profile in along-track direction with Z elevation differences of pairwise DSMs against reference Lidar DTM before (square-dot lines) and after (solid lines) image geometry correction (to be compared with elevation corrections in Figure 4.35).

values of -0.10 m, -0.32 m and 0.21 m), the difference models are reduced to mean values very close to zero. The image correction approach significantly reduced the RMS values in each scene combination with up to 34%, with highest improvement for the NB scene combination (from 0.85 m to 0.56 m).

Table 4.18: Accuracy assessment of Pléiades DSMs in open areas before and after image correction (given values are in meters).

Scene Comb.	Before image correction				After image correction			
	Mean	**Std**	σ_{MAD}	**RMSE**	**Mean**	**Std**	σ_{MAD}	**RMSE**
FB	-0.10	0.64	0.53	0.65	-0.00	0.44	0.45	0.44
FN	-0.32	0.78	0.73	0.84	-0.02	0.57	0.52	0.57
NB	0.21	0.83	0.82	0.85	0.10	0.55	0.51	0.56

4.3.4 Suitability of Coarser Resolution DSMs for Satellite Image Geometry Improvement

The proposed method is developed based on a reference high resolution terrain model from Lidar. However, in order to test the possibility of improving VHR satellite image geometry by using a coarser resolution elevation model, the open source ALOS World 3D (AW3D30) DSM was chosen. As freely available global dataset with 1" resolution (equivalent to approximately 30 m at the Equator), the AW3D30 elevation model could be considered as an appropriate reference data for the vertical accuracy assessment of the photogrammetrically derived satellite models. However, investigations with respect to its own vertical accuracy should be performed. Acquired by

the Panchromatic Remote Sensing Instrument for Stereo Mapping (PRISM) operated on the ALOS ("Daichi"), the AW3D30 was photogrammetrically developed by the Japan Aerospace Exploration Agency (JAXA) using optical imagery collected during the mission between 2006 to 2011. The geographic coordinates are referenced to the GRS80 ellipsoid (ITRF97), while elevations are with reference to the EGM96 geoid (Tadono et al. 2015). In literature, studies assessing the vertical accuracy of AW3D30 generally report values of approximatively 5 m. However, various research papers demonstrate a better performance with higher accuracies of less than 5 m (Santillan et al. 2016; Caglar et al. 2018).

In current investigation, the vertical quality of ALOS elevation model was checked against the available Lidar DTM in Allentsteig study area for the open areas, by computing the statistic parameters and by visual analysis of the colored difference model (Figure 4.40, Table 4.19). For this, the AW3D30 model was up-sampled to 1 m resolution, by applying a cubic interpolation. The results indicate a RMSE of 2.79 m with standard deviation of 2.32 m, values smaller than the accuracies of 4 m and 4.10 m reported by Takaku, Tadono, and Tsutsui (2014) and Tadono et al. (2015), respectively.

To check the possibility of using the AW3D30 elevation model as reference instead of lidar for image correction model estimation, the difference models between photogrammetric satellite DSMs and AW3D30 were computed. For this, a roughness and elevation mask was computed for removing the vegetation and build-up areas, such that only open, free areas are investigated. The resulted discrepancies are generally high (RMSE to 3 m) and consequently, the Z elevations section profile in along-track direction (Figure 4.40(c)) shows higher ondulations when compared to the ones determined by using Lidar model as reference (Figure 4.38(b)).

The statistic parameter comparison between the vertical assessments of Pléiades DEMs against Lidar DTM and AW3D30 reveals a lower performance of the AW3D30 DEM with about 2 m in standard deviation and RMSE (Table 4.19).

Table 4.19: ALOS vertical accuracy with reference to Lidar and comparison of Pléiades DEMs with respect to Lidar and ALOS in open areas (given values are in meters).

Elevation models	Lidar reference					ALOS reference				
	μ	Med	σ	σ_{MAD}	RMS	μ	Med	σ	σ_{MAD}	RMS
ALOS	0.39	0.83	2.77	2.32	2.79					
Pléiades comb.										
FB	-0.10	-0.16	0.64	0.53	0.65	0.28	0.70	2.85	2.38	2.86
FN	-0.32	-0.39	0.78	0.73	0.84	0.03	0.34	2.81	2.32	2.82
NB	0.21	0.24	0.83	0.82	0.85	0.62	1.08	2.98	2.54	3.04

To improve the vertical quality of AW3D30 model, a global LSM approach was employed for estimating a full 3D affine transformation with 12 parameters towards the Lidar reference model. This resulted in rather small improvements (0.17 m), with RMSE value decreasing from 2.79 m to 2.62 m (Table 4.20). Nevertheless, the systematic errors are still kept in the **lidar-AW3D30** difference elevation model.

Similar as before, the Pléiades DEMs and lidar were compared against the LSM AW3D30 model in open areas. From a visual inspection, after the global LSM transformation a better agreement between satellite-based DEMs and the reference
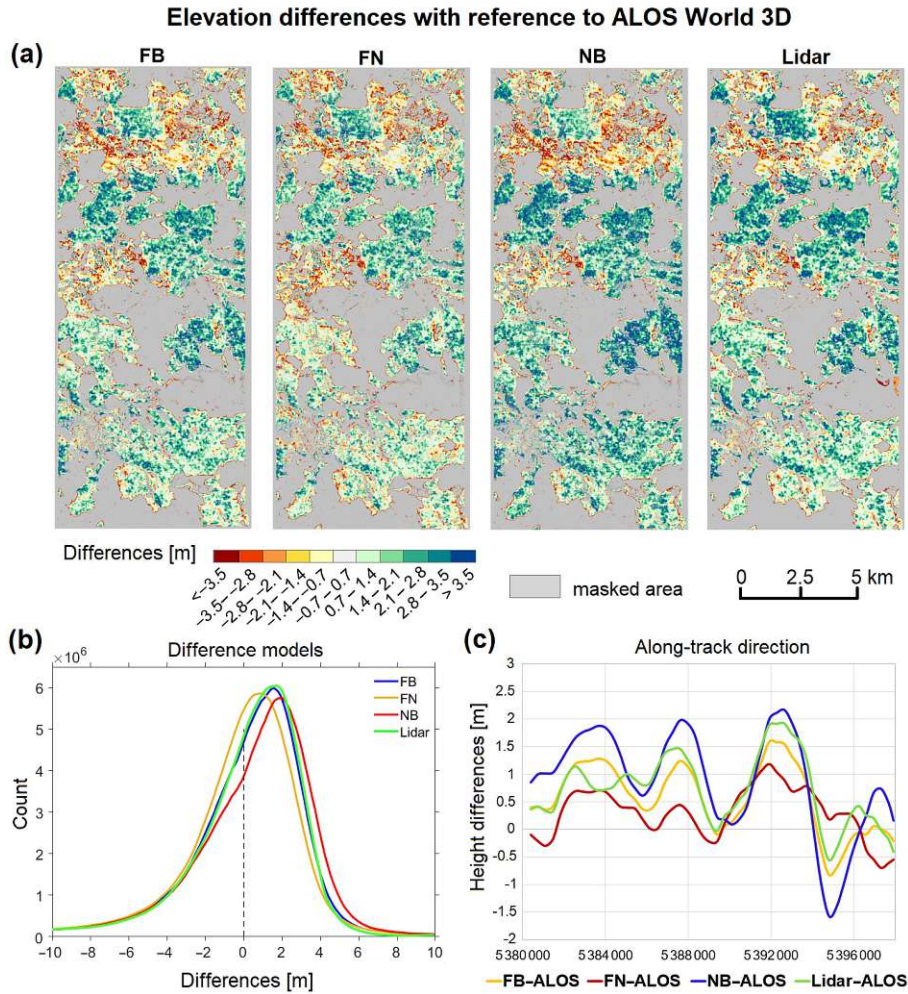
Figure 4.40: Elevation differences between Pléiades DEMs and Lidar DTM with reference to ALOS World 3D: (a) Color coded height differences models (masked areas are shown in grey); (b) Frequency distribution of height discrepancies for all combinations; (c) Section profile in along-track direction with Z elevation differences of pairwise DSMs against reference ALOS World 3D.

AW3D30 is obtained (Figure 4.41). Initially systematically shifted by ≈ 1 m, the Pléiades based heights are reduced to medians close to zero after applying LSM transformation, with the FB image combination elevation model showing better accuracy then the other two.

Since the RMSE between Pléiades DSMs and AW3D30 is ≈ 2 m the resulted elevation undulation effects in object space are higher compared to Lidar by a factor of 3 (from 0.7 m to 2.3 m). Therefore, by following a similar workflow for estimating the image corrections based on the AW3D30 model will lead to an overestimation by a factor of 3 for the correction values in image space. With its systematic errors possibly caused by sensor instability with respect to Lidar, the AW3D30 model is not accurate enough for computing the systematic errors in the Pléiades DSMs with amplitudes of maximum 0.88 m.

Table 4.20: LSM ALOS (after LSM to lidar) vertical accuracy with reference to Lidar and comparison of Pléiades DEMs with respect to Lidar and LSM ALOS in open areas (given values are in meters).

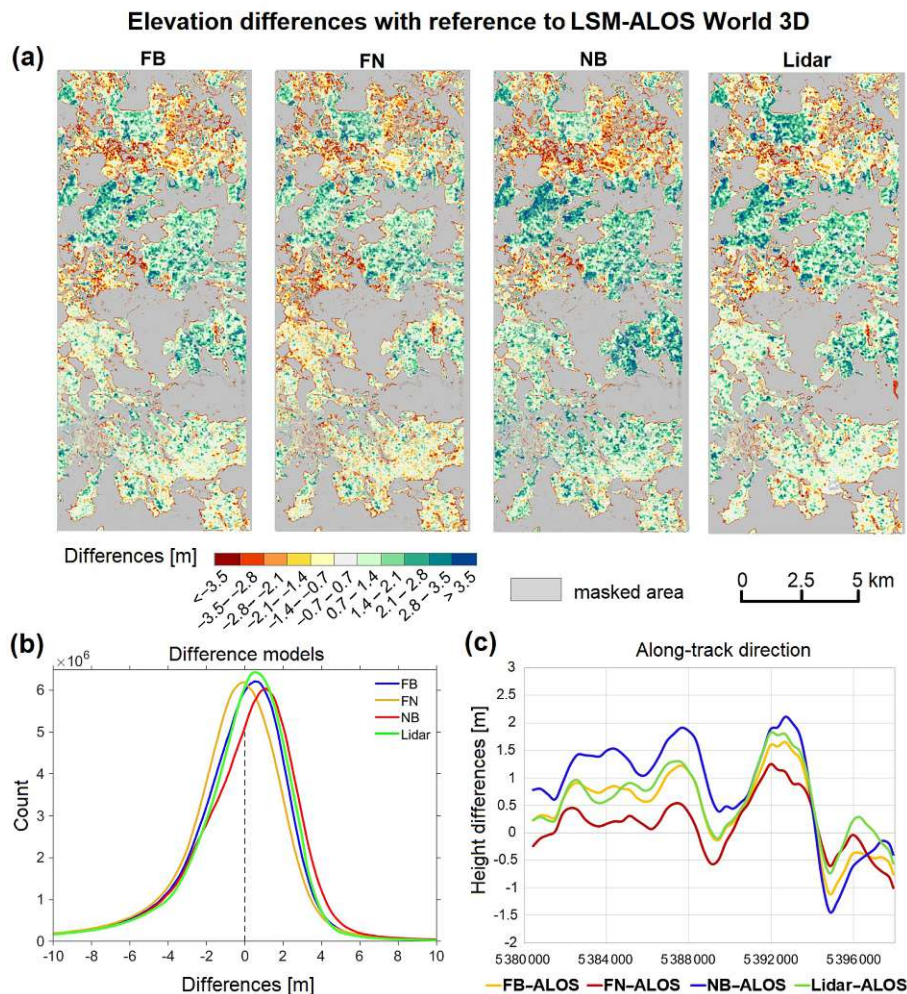| Elevation models	Lidar reference					LMS ALOS reference				
	μ	Med	σ	σMAD	RMS	μ	Med	σ	σMAD	RMS
LMS ALOS	-0.18	0.17	2.76	2.14	2.62					
Pléiades comb.										
FB	-0.10	-0.16	0.64	0.53	0.65	-0.27	0.04	2.70	2.21	2.71
FN	-0.32	-0.39	0.78	0.73	0.84	-0.54	-0.31	2.70	2.19	2.76
NB	0.21	0.24	0.83	0.82	0.85	0.05	0.42	2.82	2.36	2.82

Figure 4.41: Elevation differences between Pléiades DEMs and Lidar DTM with reference to LSM-ALOS World 3D: (a) Color coded height differences models (masked areas are shown in grey); (b) Frequency distribution of height discrepancies for all combinations; (c) Section profile in along-track direction with Z elevation differences of pairwise DSMs against reference LSM-ALOS World 3D.

4.3.5 Further Remarks

Computed with the proposed method, the resulted image correction values for the tri-stereo Pléiades satellite images (with a maximum amplitude of 0.075 pixels) align

96

very well with the values reported by Jacobsen (2018), where the Pléiades stereo scenes have systematic effects below 0.10 pixels.

The triple stereo scene allows a further verification of the image geometry correction. In Figure 4.29 the differences between the surface models of the different image pairs before image geometry correction was shown. These pairwise DSM differences after image geometry correction, $DSM_{FB} - DSM_{FN}$, $DSM_{NB} - DSM_{FB}$, and $DSM_{NB} - DSM_{FN}$, are shown in Figure 4.42. The difference is dominated by random noise and systematic patterns in height are below 14 cm. This shows the effectiveness of the proposed algorithm.
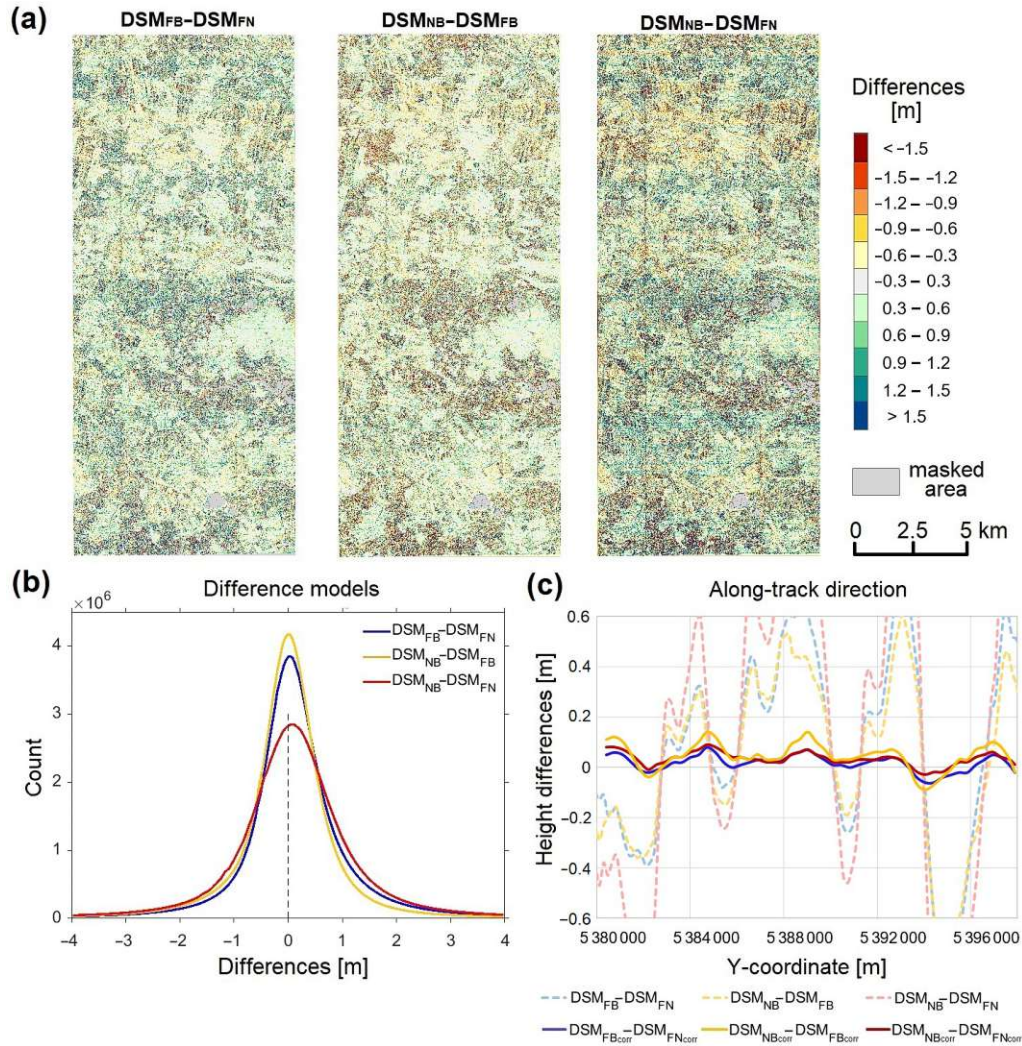

Figure 4.42: Differences between Pléiades pairwise DSMs after image geometry correction: (a) Color-coded height differences; (b) Frequency distribution of height discrepancies; (c) Section profile in along-track direction with Z elevation differences before (square-dot lines) and after (solid lines) image geometry correction (to be compared with Figure 4.29).

As presented in Table 4.18, height errors of the FB, FN, and NB error are 44 cm, 57 cm, and 56 cm, respectively, whereas the systematic errors are below 14 cm. Thus, systematic errors are clearly below the random error components. The formula for the precision of heights derived by forward intersection is: $\sigma_Z = m_B \cdot \sigma_B \cdot H/B$ (Kraus

2012), with m_B the image scale and σ_B the image measurement precision. The base height ratios B/H are given in Table 3.1. From this, the image measurement precision can be computed for each pair and amounts to 10% to 15% of the pixel size. It has to be considered, however, that the DSM interpolation also introduces some smoothing, and these values are too optimistic for the estimation of measurement precision in dense matching. This would explain the better accuracy than reported by Ressl, Brockmann, et al. (2016) for very high resolution aerial photography (GSD ¡10cm), which suggests DSM accuracy between 0.5 GSD and 1 GSD for much more favourable B/H ratios.

Comparative analyses regarding the vertical accuracy assessment have also been conducted on other Pléiades and WorldView-3 DSMs. In these datasets, the same periodic systematic vertical offsets (undulation effects in flight direction) were visible. This confirms again that attitude sensor oscillations are affecting the satellite imaging geometry, causing distortions that need to be corrected. However, these analyses need to be further investigated. The results show that not only the investigated Pléiades DSMs suffer from errors, but also the elevation model from ALOS sensor displays periodic height discrepancies when compared against Lidar DTM reference. This might be caused by the existing distortions in the ALOS prism images, also reported in (Schwind et al. 2009; Takaku and Tadono 2010).

These findings add further knowledge to the specific topic, by extending the currently existing list of optical sensors with limited orientation accuracies, as formulated by Karsten Jacobsen (n.d.). His research demonstrates that several optical sensors, among which the chinese ZiYuan-3, SPOT-5, WorldView-2, and Cartosat, have limited orientation accuracies, a fact that causes deformations in final derived elevation models.

4.4 Orthorectification of high resolution satellite imagery

4.4.1 Pléiades and World View-3 image orthorectification

With the objective of obtaining precise and high accurate ground truth data and for an efficient use of geo-information in various applications, quality images with a corrected geometry are required. Usually satellite image data are delivered with varying accuracy. However, orthorectification using only ephemeris and attitude data measured on board of the satellite provide a lower absolute accuracy in comparison with the inclusion of the GCPs in the processing workflow. For specific applications like change detection for disaster mapping, an accuracy in the sub-pixel range is necessary, thus requiring a precise orthorectification with the use of GCPs (Hoja et al. 2008).

4.4.2 Solutions for enhanced orthorectified satellite images

The Pléiades and the World View-3 satellite images were orthorectified (geometrically corrected) by removing the effects of image perspective (tilt) and relief (terrain). The entire processing workflow was performed with the Ortho Master module of the Trimble Inpho software. Through the orthorectification process the perspectiveness of each line from the pushbroom satellite images was eliminated resulting in an orthographic projection where the rays are parallel as opposed to the perspec-

tive projection where all the rays pass a common perspective center. The resulted orthophotos contain features represented in their 'true'-positions at a uniform scale in a projected coordinate system, that can be used as planimetric maps. Unlike the original satellite images, the resulted orthophotos could be then used for measuring distances (e.g. street widths), since the accurate representation of the Earth's surface allows the measurement of true distances.

A very important aspect in orthorectification is played by the elevation model, which is required to create an accurate orthophoto, as distortions in the image (due to varying distance between the sensor and different points on the ground) need to be corrected. The elevation model can be not only the DSM or the DTM of the area, but these could be replaced by a constant height plane model as well. When the stereo DSM is employed in the processing, then the true orthophoto is obtained. Compared with a normal orthophoto, a true orthophoto correctly describes not only the ground and near-ground objects, but also multi-storey buildings and engineering structures like for e.g. bridges.

The orthophotos were generated with a pixelsize of 0.70 m for Pléiades and of and 0.30 m for WorldView-3 imagery, based on the Nadir image acquisitions and using the exact Cubic Convolution rectification method (gray-value is interpolated from 4x4 pixels). Details regarding the computed orthophotos for each image are described in Table 4.21. For Stockholm, besides the photogrammetric derived DSM the open online AW3D30 surface model from ALOS World 3D (30 m spatial resolution) was used.

Table 4.21: Orthorectification of satellite imagery for each study site.

Study area	Sensor	Acq. Date	Scene size (km^2)	Elevation models for orthorectification	Product
Allentsteig	Pléiades	13-06-2017	159		
	Pléiades	20-06-2017	378	DSM (stereo)	True Orthophoto
	WV-3	08-04-2018	100		
Barcelona	WV-3	20-07-2018	100	DSM (stereo)	True Orthophoto
				DTM	Orthophoto
				Constant elev. plane	Orthophoto
Stockholm	WV-3	17-11-2017	27	DSM (stereo)	True Orthophoto
				DSM - ALOS	Orthophoto
Himalaya	Pléiades	19-10-2015	563	DSM (stereo)	True Orthophoto

The accuracy and the visual quality of the obtained orthophotos is affected by several parameters that depend on the used input data in the computation. Specifically these are: the quality and resolution of the satellite images, the interior and exterior orientation of the images and the accuracy of the digital terrain or surface model. Since the first parameters were already investigated, in the next steps the focus falls on the quality of the elevation model used in the orthorectification process.

4.4.3 Urban areas orthorectification

The geometric correction of satellite imagery over urban areas encounters difficulties caused by the residual perspective errors, which are connected to the presence of

great altimetric discontinuities due to high buildings and existent infrastructure
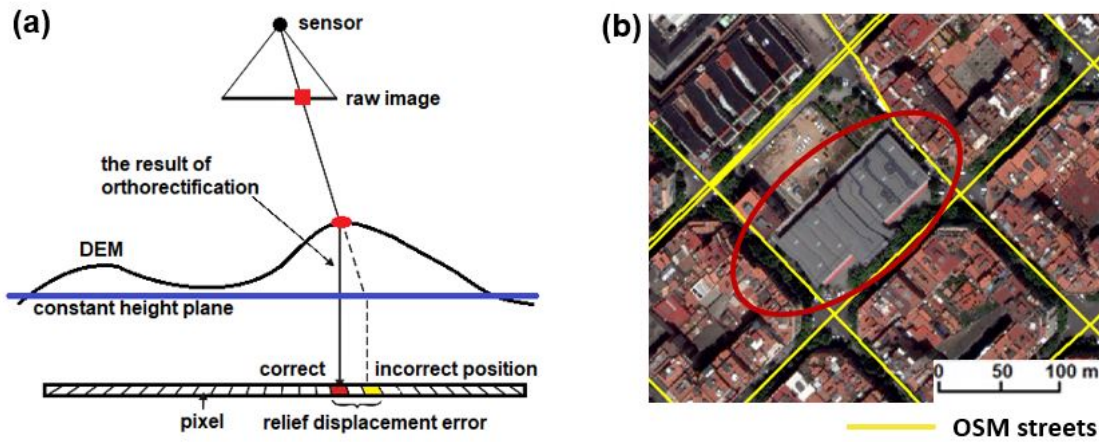(Figure 4.43).

Figure 4.43: (a) Relief displacements caused by the perspective projection; (b) Or-
thophoto detail with overlaid OSM data streets (yellow) showing warping effects on
buildings.).

4.4.3.1 Advantages and limitations of true orthophotos

The true orthophoto over Barcelona has a good geolocation, with planimetric offsets
in X and Y directions smaller than one pixel size (0.31 m). Evaluated at 23 CPs
the true orthophoto shows an overall geolocation mean of 0.33 m with a standard
deviation of 0.22 m and a RMS value of 0.41 m (Table 4.22). Here the precision
of the control point measurement could influence the results and should be con-
sidered, but in this case, the two orthophoto resolutions are very similar (0.25 m
for the reference and 0.30 m for the computed orthophoto) and the manual point
picking accuracy ranges within these values. Figure 4.44(a) shows an illustration of
a test area in the true orthophoto, that has a good geoposition with objects recti-
fied to their correct locations, but shows distortion effects and artefacts along roof
boundaries, caused mainly by the high roughness of the photogrammetric derived
DSM. Moreover, buildings obscure the objects close to them, since their walls can
be thought of as a rapid change in elevation. The orthophoto does not detect these
occluded areas and creates instead a "double projection", known as "ghost images"
or "double mapping effect". When a building is orthorectified, then it will get the
original position in the rectified image, but it will also leave a "copy" of the building
on the terrain. This fact is illustrated in the right image of Figure 4.44(a), where the
roof of the buildings are double projected. During reprojection, rays are projected
back to both the occluded area and the obscuring object (building), therefore the
"false" image data is rectified in the occluded areas.

4.4.3.2 Constant height plane- and DTM-based orthophoto

In ideal cases, the true orthophoto would be the best solution for further analyses,
but for Barcelona study area, this has some limitations (Figure 4.44(a)). For en-
hancing the quality of the orthophotos, by reducing the distortion, double mapping
effects and artefacts, the photogrammetric DSM was replaced by different height

models. Specifically, a plane with a constant height and the corresponding DTM were considered. The newly obtained orthophotos have the disadvantage of not handling sharp changes in elevation because the terrain model does not include any objects on it, i.e. buildings and vegetation. Hence, they will contain relief displacements, (i.e. the position of an object in the image is shifted) caused by the height above the terrain (Figure 4.43(a)). As a constant height, the median value of the photogrammetric DSM elevations of 82 m was considered.

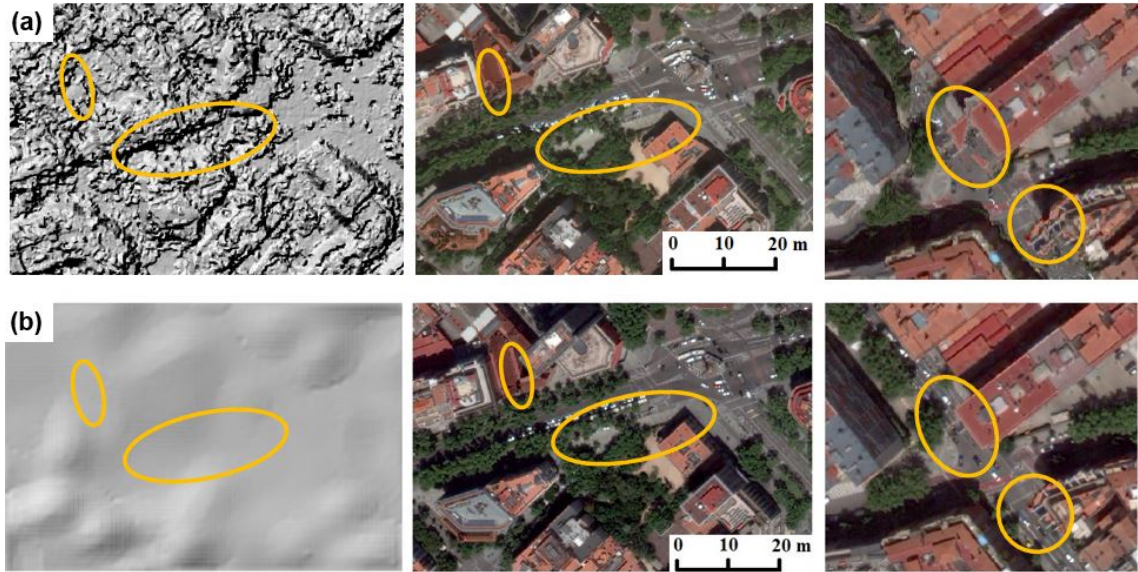For a comparative overview, Figure 4.44 shows the DSM- and DTM-based orthophoto for the same area.

Figure 4.44: (a) Artefacts and distortion effects in the DSM-based orthophoto (center and right) caused by the high roughness of the image-matched DSM (left); (b) DTM (left) and DTM-based orthophoto (center and right).

The DTM-based orthophoto shows similar statistics with the true orthophoto regarding the planimetric accuracy evaluated at the CPs (Table 4.22). This fact is explained by the location of the points on the ground surface, where the terrain level is being correctly reprojected. The DTM-based orthophoto has a good fit with the 0.25 m reference orthophoto, but still some warping effects are visible especially for buildings. In contrast, the orthophoto build by using the constant elevation plane shows higher displacements, with an overall mean of 11.24 m and 10.49 m standard deviation. Even if due to the constant elevation no distortion and warping effects are visible, the displacement error can reach up to 37 m at higher altitudes. This product is comparable with the original OR2A satellite images (map-projected products using a DEM with an elevation of 132.58 m and no topographic relief with respect to the reference ellipsoid), but in this case, the elevation plane was data-driven. Moreover, the geolocation accuracy is higher for the 82 m constant elevation plane-based orthophoto than for the initial images based on the 132.58 m elevation plane. While the mean of the overall planimetric displacement for the forward OR2A image is 19.10 m (Table 4.3), the derived orthophoto shows a mean shift of only 11.24 m.

Since buildings are not modeled in the DTM nor in the constant height plane, they are leaning away from the image nadir point and their facades are visible in the computed orthophoto, occluding the surrounding areas (Figure 4.45). Only

objects that are at the terrain level are correctly reprojected. For Barcelona city, the geoposition difference between the 2 orthophotos is < 1 pixel (0.3 m) at ground level. Their planimetric position fits very well with the reference OSM data.
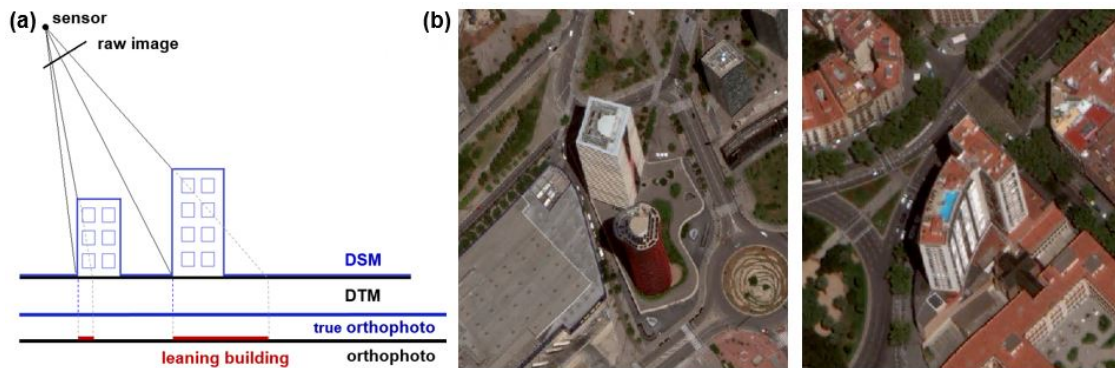

Figure 4.45: Leaning-buildings effect: (a) Illustration of central perspective projection with true and normal orthophoto mapping; (b) Tall buildings with shadows leaning over streets hiding information.

Table 4.22: Evaluation on 23 GCPs of the planimetric accuracy for Barcelona orthophotos (given values are in meters).

Barcelona Orthophoto (OP)	Planimetric displacement						Overall displacement		
	in X-direction			in Y-direction			XY		
	μ	σ	RMS	μ	σ	RMS	μ	σ	RMS
True OP	-0.002	0.23	0.22	0.21	0.27	0.32	0.33	0.22	0.41
DTM-based OP	0.12	0.21	0.23	0.25	0.30	0.34	0.34	0.23	0.41
Cst. el. OP	-3.42	10.26	10.59	-5.52	13.94	10.92	11.24	10.49	15.21

4.5 Classification of satellite derived point clouds

4.5.1 Radiometric calibration of satellite imagery

As a pre-processing step, an optical radiometric calibration of the images is required before any comparison between them is performed. Depending on the season, the atmospheric conditions, the sun's azimuth and elevation, the energy that satellite sensors record is different from the actual energy emitted or reflected from a surface on the ground. Since the value recorded in each pixel includes not only the reflected radiation from the surface but also the radiation scattered and emitted by the atmosphere, the pixel values, i.e. Digital Numbers (DN), need to be corrected.

The tri-stereo Pléiades images were provided as primary product at sensor processing level, corrected only from on-board distortions such as viewing directions and high frequency attitude variations (Panem et al. 2012) However, this was only partly successful, as it could be seen in Section 4.3. For all the three images the optical radiometric calibration was performed using the open source software Orfeo Tool Box (Inglada et al. 2009). The pixel values were calibrated by the influence of the following parameters: sensor gain, spectral response, solar illumination, optical

thickness of the atmosphere, atmospheric pressure, water vapour and ozone amount, as well as the composition and amount of aerosol gasses.

In contrast, the WorldView-3 images were delivered at OR2A processing level with relative radiometrically-corrected image pixels. The relative radiometric calibration includes a dark offset subtraction and a non-uniformity correction (e.g. detector-to-detector relative gain) which is performed on the raw data during the early stages of product generation. The absolute radiometric calibration for each WorldView-3 image was computed. This was done in two steps: (1) conversion from image pixels (DN, raw data from the sensor) to top-of-atmosphere spectral radiance, based on the rescaling factors provided in the metadata file, and (2) conversion from top-of-atmosphere spectral radiance to top-of-atmosphere reflectance. This requires additional information like: Earth-Sun distance, solar zenith angle and solar spectral irradiance. The calculations were performed for each image (and independently for each band), by using the equations and parameters found in the technical sensor description (Kuester 2016). Usually, the optical radiometric calibration step is necessary before making any physical interpretation of the pixel values. In particular, this processing is mandatory before performing any comparison of pixel spectrum between several images from the same sensor. In this case, this pre-processing step could be omitted since it did not change the geometric quality of the images.

Optical radiometric calibration is important in training the classifier without dependence to the atmospheric conditions at the acquisition time, making it valid for a wider range of applications and not particular to one study case only. The optical calibration of the images is definitely required when investigating the transferability performance of the classifier across different acquisition times. The cross-time performance is checked by applying the classifier trained on Pléiades first acquisition to the Pléiades second acquisition.

4.5.2 Experiment design

Once the satellite image-derived 3D point clouds are generated, a supervised tree based classification algorithm using machine learning is applied for labelling the points into the five following classes: ground, trees, roads, buildings, and vehicles. In this case, the class ground comprises not only the bare-soil but also the covered ground, such as the agricultural- or grass-lands and all points not being included in the other classes. The focus falls on these five specific classes because they satisfy the typical requirements taken from an industry related project, with an interesting aspect that some classes occur rarely, while others more often.

The classification algorithm was applied on the point clouds derived for the Allentsteig study area corresponding to the three different acquisitions: Pléiades first (13-06-2017), Pléiades second (20-06-2017), and WorldView-3 (08-04-2018). For a comparative analysis in the further investigations, an interest area of $24 km^2$ from the overlapping of these three datasets was considered (Figure 3.3). This hilly region is characterized by forests, open areas and in the southern part more villages are found. To obtain training and testing data the three point clouds of the entire study area were manually classified using the application TerraScan from Terrasolid software. With its versatile visualisation options, the application allowed the manually labelling of the 3D point clouds into ground, trees, roads, buildings, and vehicles. Additionally, the corresponding orthophotos for each data acquisition were used in

the thorough visual analysis.

In a first step, additional geometric features were computed for each 3D point in the matched cloud. For this, in the spatial query of a point neighbourhood, an infinite cylinder with 7 m search radius was considered. The local tangent plane estimated by computing the best fitting plane for the nearest points and its corresponding normal vectors together with the standard deviation of the fit ($NormalSigma0$) are used as additional descriptors of the points. The computed features include the normal vector components ($NormalX$, $NormalY$ and $NormalZ$), features derived from the structure tensor, vertical point distribution, and surface roughness. These additional attributes describe the point distribution and are required for the separability of classes. The *linearity*, *planarity* and *omnivariance* features were computed from the structure tensor eigenvalues. The first two have values between zero for non-linear/non-planar objects and one for linear/planar point distributions, respectively. The *omnivariance* feature gives information about the volumetric distribution of points, while the $EchoRatio$, $ZRange$ and $ZRank$ describe the vertical point distribution, the maximum height difference between neighbouring points, and the rank of the point corresponding to its height, respectively. Besides geometry, point clouds derived from image matching have radiometric information (R, G, B, nIR), which comes from the image texture after applying the radiometric calibration. Therefore, in addition to the hand-crafted 3D geometric features, the RGB colour information in each point was used for training the classifier. For the model learning process a total of 17 features were used: *Red*, *Green*, *Blue*, *EchoNumber*, *Numberofechos*, *Amplitude*, $NormalX$, $NormalY$, $NormalZ$, $Normalsigma0$, *linearity*, *planarity*, *omnivariance*, $EchoRatio$, $NormalizedZ$, $dZRange$, and $dZRank$ (Waldhauser et al. 2014; Bachhofner et al. 2020).

The experiment is designed to characterize the repeatability of the semantic segmentation for satellite derived 3D point clouds. Thus the same area is studied, and also the same areas are used for training the classifier and testing the model. The setup of the experiment is to compare the classification results over the same areas for the different acquisitions. For evaluating the precision and performance of the classifier, the variation of the results for three scenarios is analysed. For this, firstly the point clouds generated by image matching were divided into five line-patches each with the size of 0.4 x 7 km and a new model was repetitively trained and validated by employing distinct combinations for the test and training data, like shown in Figure 4.46. Therefore, the proposed model is trained on 40% of the data, while 60% is being used for validation. To obtain more stable and robust results, the experiments were repeated 5 times for each scenario and data acquisition and classification metrics are associated with standard deviations.

The classification tree is trained with the following hyperparameters: 0.00001 complexity factor, minimum 20 observations existent in a node (for splitting), at least 7 observations for each leaf node, a maximum depth of 30 for the tree, 5 competitor splits, and 5 surrogate splits (Bachhofner et al. 2020). Finally, to estimate how accurately the predictive model is, it was tested on the validation dataset, containing labelled points. The same processing chain comprising the machine learning supervised classification algorithm was applied independently in each scenario and for each of the three datasets: Pléiades first, second and WorldView-3 acquisitions.

The distributions of the number of points per class in each scenario and for each image acquisition are given in Table 4.23 as percentage values. The number of
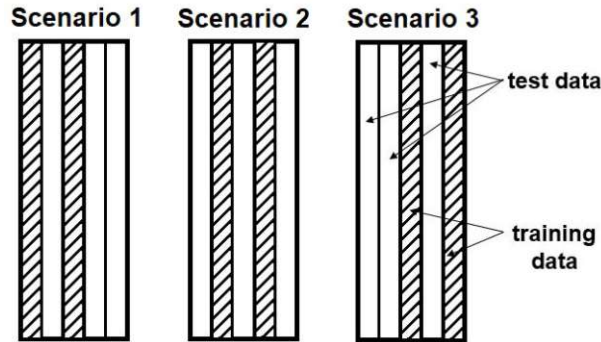
Figure 4.46: Overview of the different scenarios with the selections for the training and test data.

instances included in classes Buildings, Roads and Vehicles is far less than the ones included in Ground and Trees.

The distribution of the classes is similar between the acquisitions and also between the scenarios. The largest variation in absolute numbers is found between the two Pléiades point clouds on the one hand and the WorldView 3 point cloud on the other hand for the classes ground and trees. For the classes, roads and buildings the variation between the scenarios is larger than the variation caused by acquisition time.

Table 4.23: Class distribution for each scenario and image acquisition (values given in percentages).

| Class | Percentage of points / Class | | | | | |
| | Pléiades (1st) acq. | | Pléiades (2nd) acq. | | WorldView-3 | |
	Train.	Test	Train.	Test	Train.	Test
Scenario 1						
Ground	52.748	60.257	52.887	60.376	66.474	71.664
Trees	45.351	38.119	45.250	38.064	31.682	26.708
Roads	1.641	1.430	1.610	1.366	1.544	1.423
Buildings	0.257	0.189	0.250	0.191	0.297	0.203
Vehicles	0.003	0.004	0.002	0.002	0.002	0.002
Scenario 2						
Ground	57.611	57.136	57.726	57.272	71.367	68.445
Trees	40.569	41.194	40.516	41.106	26.854	29.889
Roads	1.542	1.492	1.479	1.449	1.482	1.462
Buildings	0.276	0.174	0.279	0.169	0.295	0.201
Vehicles	0.003	0.004	0.001	0.003	0.001	0.003
Scenario 3						
Ground	58.007	56.868	58.166	56.975	69.159	69.971
Trees	40.576	41.185	40.492	41.119	29.393	28.137
Roads	1.232	1.704	1.163	1.665	1.248	1.622
Buildings	0.179	0.241	0.176	0.241	0.199	0.268
Vehicles	0.006	0.002	0.003	0.001	0.002	0.002

The chosen classes have a decreasing occurance probability. In such situations, the overall accuracy measure alone is inappropriate, since the big number of examples

from the majority classes overwhelms the number of examples in the minority classes. Therefore, to counter class imbalance metrics such as (average) Precision, (average) Recall, and F1-score were employed for evaluating the prediction performance of the classifier.

4.5.3 Repeatability and transferability of Supervised classification

During classification, an object class label is assigned to each point, resulting in a 3D labelled point cloud. From a visual inspection, Figure 4.47 depicts a build-up area with the reference and classification results from all three acquisitions.
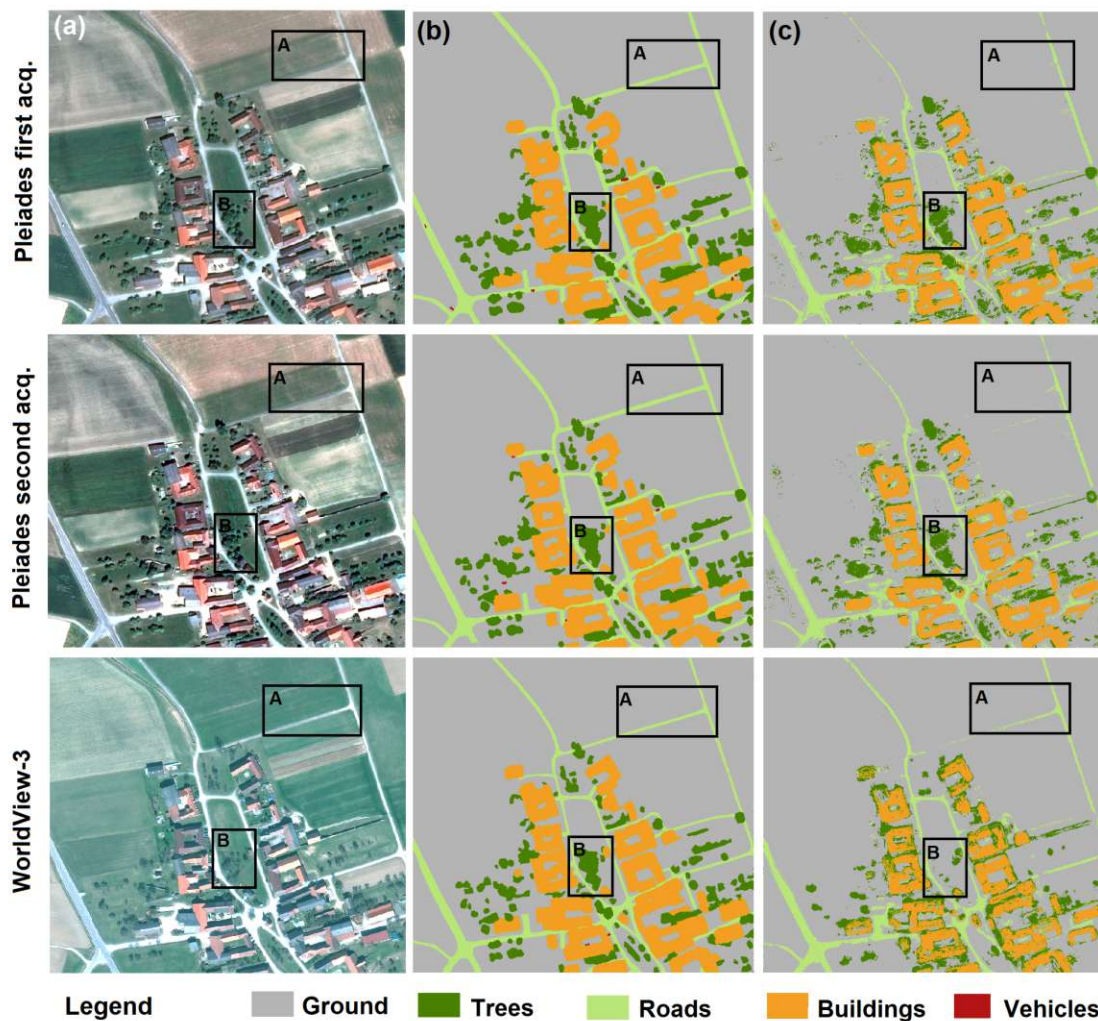

Figure 4.47: Top detail-view of Kleinschönau village (a) Pleiades and WorldView-3 ortho-rectified images; (b) the reference labelled 3D point cloud generated by image matching; (c) the classification result.

Overall, the classifier identifies ground, trees, roads, and buildings classes. The only exemption are points belonging to vehicles, which were incorrectly classified as ground or trees. From a comparative point of view, no significant differences between the results for the two Pléiades acquisitions can be observed. In contrast, for WorldView-3, some distinctions are visible. For instance, as shown in the marked area A in Figure 4.47, in the WorldView-3 acquisition points are correctly classified

as road, whereas in the Pléiades data they are wrongly labelled as ground. This is mainly because of the low Pléiades color contrast in this area. Road points have very similar geometric features with ground points and only color information is the distinctive attribute used for classification. Another notable difference appears in the marked area B in Figure 4.47. Whereas in the Pléiades acquisitions points are correctly identified as trees, in the World View-3 data only few of them are classified as trees and the rest as ground.

This situation is caused by the different acquisition times: the two Pléiades in June 2017 (leaf-on conditions) and the WorldView-3 in April next year (leaf-off conditions). The leafless appearance in the WorldView-3 images caused difficulties in the 3D reconstruction, resulting in a point cloud with no or reduced height information. Hence, the points were misclassified as ground. The confusion matrices are built for each scenario and iteration of the datasets and the achieved average results are shown in Table 4.24. In all tests, very high values for the accuracy (between 93.45% and 95.81%) are achieved, but this alone is not a reliable performance metric to use, because of the imbalanced dataset.

Table 4.24: Comparison of evaluation metrics for all scenarios. In bold are the highest and the lowest values.

Evaluation Metrics (%)		Pléiades (1st acq.)	Pléiades (2nd acq.)	WorldView-3
Average Accuracy	Scenario 1	94.58	94.51	**95.81**
	Scenario 2	94.82	95.02	95.54
	Scenario 3	**93.45**	94.12	95.50
Average Precision	Scenario 1	60.17	60.11	59.74
	Scenario 2	62.21	**62.81**	60.18
	Scenario 3	60.04	59.53	**57.47**
Average Recall	Scenario 1	68.84	67.15	68.23
	Scenario 2	**65.58**	65.81	67.03
	Scenario 3	66.78	68.12	**69.28**
Average F1 score	Scenario 1	64.18	63.54	63.72
	Scenario 2	63.75	**64.35**	63.39
	Scenario 3	63.05	63.53	**63.03**

Overall, the results are very consistent for the shown metrics, which integrate over the results of all the classes, independent of sensor and acquisition date. For the Pléiades data the results suggest that the scenarios, i.e. the choice of test data region, has more impact than the acquisition date. The WorldView results deviate stronger from the Pléiades data. At this stage it cannot be concluded if this is an effect of GSD, geometric accuracy, or acquisition time.

The qualitative results show that WorldView-3 classification metrics vary when compared with the two Pléiades acquisitions. For instance, the highest accuracy (95.81%) and average recall (69.28%) are reached in scenarios 1 and 3, respectively. Even higher variations between WorldView-3 and Pléiades can be observed at the individual class level (Figure 4.48). For the ground and road class both recall and F1-measure achieve better values in the WorldView-3 model. This confirms the visual analysis given above. The average recall raises with 3% for ground and with 8% for roads compared to Pleiades 1st acquisition and with 5% compared to Pleiades 2nd acquisition (for Scenario 1).
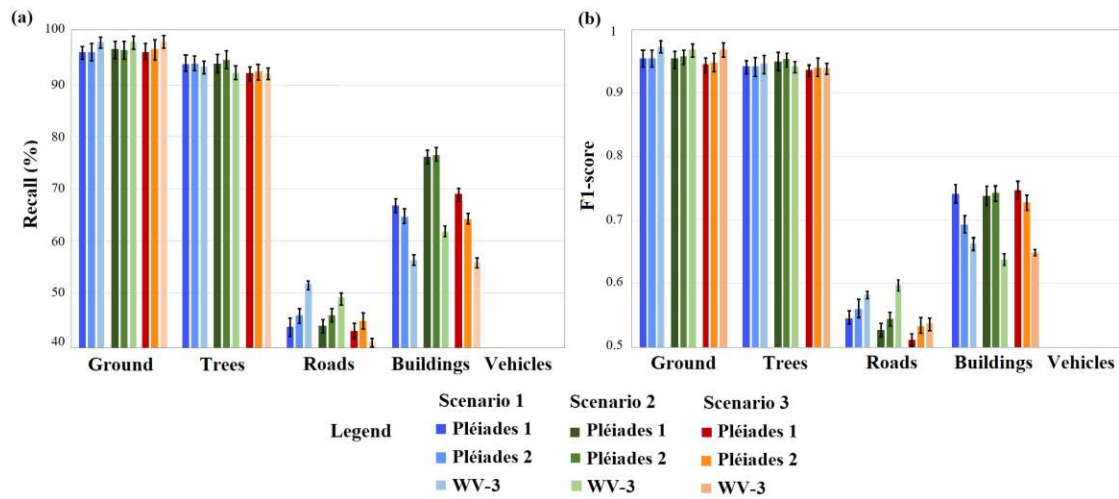
Figure 4.48: Comparison of average classification metrics with standard deviations (below 3%) for each class in all three scenarios: (a) recall and (b) F1-score.

However, the statistics for buildings drop from the Pléiades to the WorldView results by a maximum of 15% (Figure 4.48). Given the clearer appearance of buildings in Figure 4.47, this is unexpected. This can be explained, on one hand, by the leafless trees in the WorldView-3 images, which have similar radiometric properties as buildings and are therefore miss-classified as buildings. On the other hand, in all three acquisitions many points surrounding buildings are miss-classified as trees (Figure 4.49c). These particular areas are raising difficulties for the classifier, because the two classes show similar geometric features. Even colour information does not make the difference here, due to shadowing effects.

At the individual class level, the unbalanced data makes the classifier biased toward the ground and trees classes, while dwarfing the buildings, roads, and vehicles. For the vehicle class, the precision, recall, and F1-score are 0.00% in all scenarios. This is because they were no points correctly classified as vehicles (no true positives). For the class roads the highest variation of 7% F1-score is obtained between Scenario 2 and Scenario 3 when using the WorldView-3 data.

In the context of this investigation, repeatability refers to the agreement between the independent results, obtained by applying the same classifier to the same data within short interval of times. This was possible for the Pléiades acquisitions due to the temporal difference of only 7 days. Since land cover might change and trees appearance in images can differ, buildings and roads were considered as stable classes in time, with well-defined shapes and boundaries. Whereas the accuracy changes are small (below 0.15%) for both buildings and roads classes, the F1-score varies between -2.1 and 2.6 in scenario 3 (Figure 4.50). Besides the quality of the classifier, the repeatability results suggest also the stability of the Pléiades sensor itself, with respect to the spectral geometric precision and geolocation accuracy.

The transferability performance of a classifier across different acquisition times is an important question. How does the classifier trained on a dataset perform on a new dataset for the same area, but from a different acquisition time? For testing the cross-time performance, the classifier trained on the Pleiades 1st acquisition was applied on the test set of Pleiades 2nd acquisition and vice versa and the prediction performance was compared to the one achieved when training and test set are from
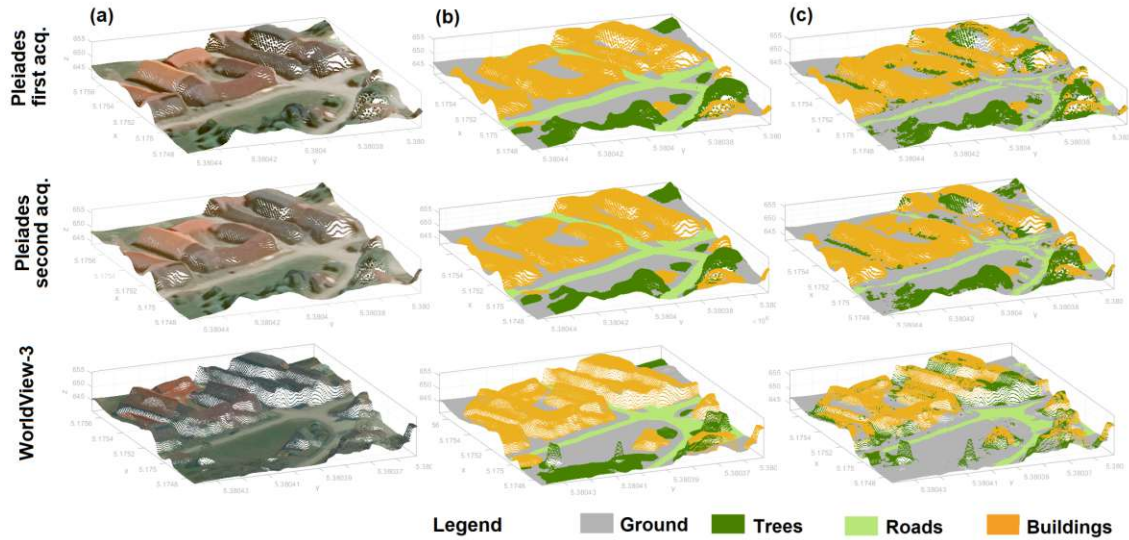
Figure 4.49: Detailed 3D view of point cloud classification results: (a) RGB point cloud; (b) reference labelled 3D point cloud; (c) classification result.
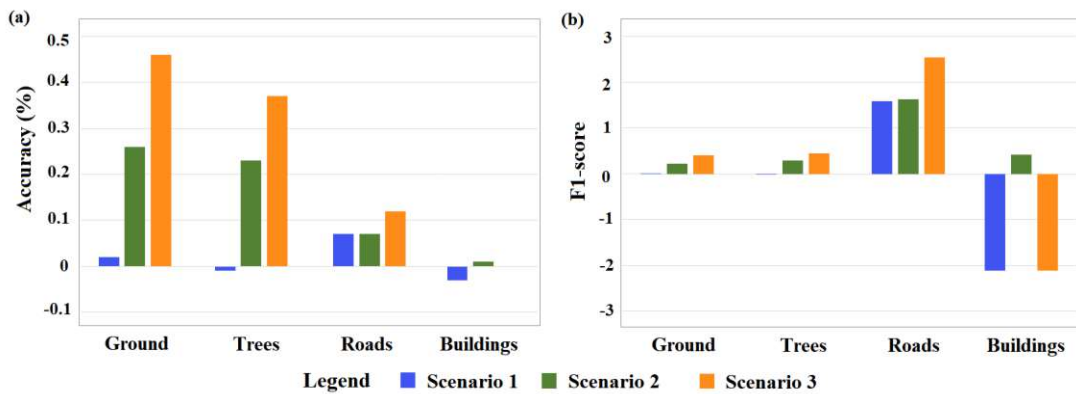

Figure 4.50: Comparison of classification metrics changes for each class (except vehicles) in all three scenarios: (a) accuracy and (b) F1-score changes between the two Pléiades acquisitions.

the same acquisition time. The results are shown in Table 4.25, for the F1-score in each scenario.

The F1-measures do not change significantly when the classifier is trained with data from another acquisition time. That proves a good transferability in time, which allows utilising the same classifier for future acquisition datasets for the same area. Overall, the evaluation metrics show a slightly higher variation between the three scenarios when compared with the changes between the two Pléiades acquisitions (Figure 4.47). Hence, the selection of the training data for building the classifier has a higher influence on the performance, than the different acquisition times.

Another critical factor that influences the performance of the classifier is the correctness of the manual ground truth annotations in all three datasets. Due to the smoothed geometric appearance of the point clouds generated by image matching, with unclear object contours labelling is a challenging task and some uncertainty may occur.

Table 4.25: F1-score result for the cross-time performance.

		Tested on	
		Pléiades 1st acq.	Pléiades 2nd acq.
		Scenario 1	
	Pléiades 1st acq.	64,26	63,26
	Pléiades 2nd acq.	60,26	63,49
		Scenario 2	
Trained on	Pléiades 1st acq.	63,87	61,87
	Pléiades 2nd acq.	61,82	64,32
		Scenario 3	
	Pléiades 1st acq.	63,19	62,92
	Pléiades 2nd acq.	60,97	63,56

Besides the variation of the classification metrics between Pléiades and WorldView-3, the higher GSD of the latter leads to higher computation times. With the same hardware configuration, an 2 x AMD EPYC 7302, 3GHz, 16-Core Processors, 512 GB RAM memory, (64-bit operating system), the processing times are approximately three times higher for World View-3 point clouds compared to Pléiades (Table 4.26). Therefore, in future work, Graphic Processing Unit (GPU) programming and parallelization schemes will have be to exploited to further reduce the computing time. It will also be necessary to include strategies to counter the class imbalanced (data augmentation, resampling, or oversampling), for improving the performance of the model. Another direction of further research would be the use of the classified point clouds for 3D object reconstruction and modelling.

Table 4.26: Processing times for classification.

Processing Step	**Processing Time / Scenario**		
	Pléiades 1st acq.	Pléiades 2nd acq.	WorldView-3
Feature computation	9h 42'	9h 11'	31h 40'
Train model	3h 17'	2h 54'	10h 46'
Apply model	28'	31'	1h 20'
Total time/iteration	13h 27'	12h 36'	43h 46'
Total (5 iterations train and apply)	~18h 45'	~17h 5'	~60h 30'

The agreement between the outcomes for the two Pléiades datasets with very low variations, assure that the reported quality measures did not result by chance. Moreover, it confirms the stability of the sensor itself, in terms of spectral geometric precision and geolocation accuracy. Additionally, the behaviour of the classifier when using a different sensor (WorldView-3) with a higher resolution and a distinct acquisition geometry is shown. From these investigations, a factor of three could be established between WorldView-3 and Pléiades data, with respect to point cloud densities (the total number of points) and processing times. This lead to a moderate improvement for some classes (ground, roads). For the class building the higher resolution and geometric sharpness provided by WorldView-3 did not lead to higher quality. In summary, it can be stated that the adopted tree based supervised learning classifier is stable, showing a good performance with overall accuracies and providing

a high potential in VHR satellite tri-stereo point clouds scene classification.

4.5.4 Classification performance and comparison with FCN-8s

An additional comparison investigation of the classification was performed for the Pléiades satellite-derived point clouds. Two different approaches were adopted for the three-dimensional classification:

1. decision tree trained on geometric features only
2. decision tree trained on geometric and color features in each point

Moreover, the Siemens "ReKlaSat 3D" project partner (*https://new.siemens.com/ at/de.html*) applied a two-dimensional classification approach, represented by a fully convolutional neural network (FCN-8s) trained on the corresponding 2D Pléiades orthophoto. This comprises the following steps: image classification, object detection, and semantic segmentation, which finally lead to a single label per image pixel (J. Long et al. 2015).

A visual analysis of the 3D machine learning classifications shows better results for the approach that employs both geometry and color information in each point. When using geometry only, buildings are miss-classified as trees and roads are not identified, mainly because of their similar geometric properties with trees and ground, respectively (Figure 4.51).
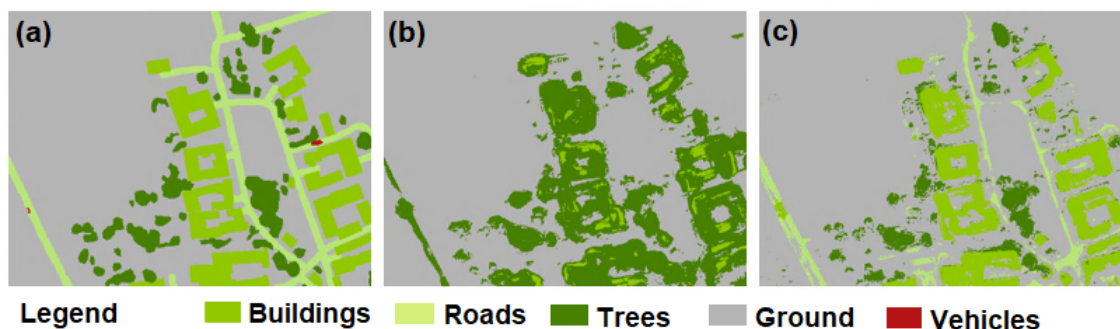

Figure 4.51: Top view of satellite-derived point cloud coloured by: (a) reference classification; (b) machine learning classification trained on geometric features only; (c) machine learning classification trained on geometric and color features.

The classification result from the FCN-8s 2D approach (Figure 4.52(d)) shows a clearer appearance and class identification, without the "salt and pepper effect" visible in the 3D classification of point clouds. This is because the 2D image semantic segmentation assigns a class to each pixel, making use of the color information and because the individual elements (points or pixels) are classified independently in the decision tree, but not independent from neighbours in the convolutional network.

The statistic classification metrics show that the fully convolutional neural network outperformes the supervised classification with an overall accuracy of 96.11%, average precision of 62.43%, average recall of 61.15% and average F1-score of 59.12% Adding color to the geometry information in the supervised 3D classification leads to a higher stability of the training process, showed by the increase of the overall performance metrics (Table 4.27).

With the color information the performance metrics at the individual class level enhance, as well. For example, the precision per class improves from 90.68% to
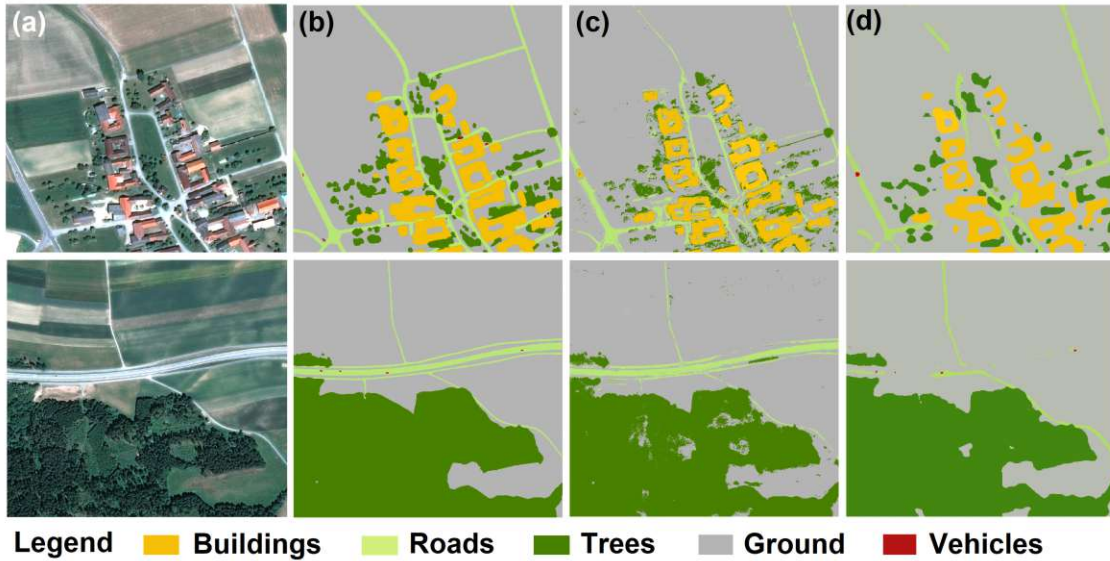
Figure 4.52: Comparison view of classification results for two areas: (a) Pléiades orthorectified images; (b) reference labelled 3D point cloud; (c) supervised decision tree classification result and (d) FCN-8s classification result.

Table 4.27: Overall quantitative comparison of the supervised decision tree (geometry only, geometry + colors) and FCN-8s (values in bold format are the highest numbers for the corresponding metrics).

Models	Overall Accuracy %	Avg. Precision %	Avg. Recall %	Avg. F1-score %
Decision Tree (3D)				
Geometry	89.10	43.89	38.73	39.54
Geometry, Colors	93.18	61.03	58.72	58.96
FCN-8s (2D)				
Colors	**96.11**	**62.43**	**61.15**	**59.12**

93.87% for ground, from 0.0% to 49.56% for roads, from 86.10% to 93.89% for trees, and from 42.67% to 67.76% for buildings. This value of 67.76% for buildings is the best performance among the three approaches (Table 4.28).

The additional color information also leads to the best recall for roads and the best F1-score for buildings. Even if the FCN-8s achieves the best results among the three approaches, the performance of the decision tree when is simultaneously trained with geometry and color information is only 1–3% lower. Moreover, for the individual classes, the decision tree outperforms FCN-8s in three cases: first, the decision tree has a higher precision for predicting buildings; second, has a higher recall for predicting roads; and third, has a higher F1-score for buildings. It is to be mentioned that none of the approaches was able to identify vehicles. These results were published in the "ReKlaSat 3D" project joint research paper (Bachhofner et al. 2020).

Finally, it is important to realize that the most important factor in determining whether a classification application will be successful or not, relies not in which the machine learning method is chosen to be applied, but in how it is used: the amount and the quality of the training data, the way the data is pre-processed and

Table 4.28: Comparison of classification metrics per class for the decision tree and FCN-8s approaches (Values in bold format are the highest numbers per class for the corresponding metric).

Model	Ground %	Roads %	Buildings %	Trees %	Vehicles %
Precision per Class					
Decision Tree (3D)					
Geometry	90.68	00.00	42.67	86.10	00.00
Geometry, Colors	93.87	49.56	**67.76**	93.89	00.00
FCN-8s (2D)					
Color	**96.09**	**63.85**	54.25	**97.95**	00.00
Recall per Class					
Decision Tree (3D)					
Geometry	93.91	00.00	11.72	88.02	00.00
Geometry, Colors	95.81	**24.76**	79.13	93.89	00.00
FCN-8s (2D)					
Color	**98.02**	22.89	**86.45**	**98.22**	00.00
F1-score per Class					
Decision Tree (3D)					
Coordinates	92.27	00.00	18.39	87.76	00.00
Coordinates, Colors	94.83	33.02	**73.01**	93.93	00.00
FCN-8s (2D)					
Color	**97.13**	**33.70**	66.66	**98.02**	00.00

represented, the way the results are interpreted and applied, and so on. There is no clear-cut method that tells in advance which machine learning method and which representation of the training data will lead to the best results. In practice, machine learning is a wide pragmatic art that involves a lot of experimentation and testing, and often the best solution can only be found by trial and error.

4.6 Potential of Pléiades and WorldView-3 tri-stereo DSMs to represent heights of small isolated objects

The standard method of DSM generation from stereo-pairs or triples is dense image matching using global or semi-global optimization. Caused by the smoothness constraint of dense image matching (Stentoumis et al. 2015) the heights of small individual objects may not be reconstructed. Hence, the corresponding 3D points will not have higher elevations compared to their surroundings. Based on this hypothesis, the capability of dense image matching when evaluating the height of small individual, i.e. detached, objects, was investigated. The object's height compared with its surrounding terrain (in the following simply referred to as height) was investigated with respect to the following parameters: building length and width, vehicle length and width, and tree crown diameter. The workflow for this analysis comprise the following processing steps: satellite image block orientation (in Section 4.1) using RPCs and a set of GCPs, 3D reconstruction (in Section 4.2.1), monoscopic measurements in all three images, computation of the reference individual object's height (by subtracting the correspondent elevations from the lidar DTM and in the last stage, comparison of the automatically reconstructed heights with the reference

values for different object classes.

4.6.1 Manual reference measurements

The manual measurement of the 3D points of interest corresponding for isolated objects was performed after image orientation. For each object, the highest point was selected, such as points on the building's ridge, on car's roof and tree crown centre (approximation of the tree top). For the 3D restitution, the points were manually monoscopically measured in all three oriented images: Forward, Nadir and Backward, in a multi-aerial view mode. The mean-square error of the Z object coordinate is given by the following formula (Kraus 2012):

$$\sigma_Z = \sqrt{2} \cdot m_B \cdot \sigma_B \cdot \frac{Z}{B} \qquad (4.2)$$

with σ_Z the object elevation accuracy, σ_B the accuracy of image measurement ($1/3^{rd}$ of a pixel), Z the altitude of the satellite orbit, B the base and m_B the satellite image-scale number given by the Z/c ratio, where c is the focal length of the optical system. Due to large heights, the special case of parallelism between object and image plane was considered; hence a single scale number for the whole satellite image was used. Taking into account these parameters, the estimated accuracy of the Z object coordinates is 1.36 m and 0.31 m for the Pléiades and WorldView-3 images, respectively. These results suggest a minimum object height of 1 m as a reference height that guarantees a reasonable analysis. Since the smallest investigated cars have around 1.5 m height and buildings and single trees usually have more than 2 m, the estimated elevation accuracy does not significantly influence the investigations. In a final step, the reference object heights are computed by using the measured Z-coordinates and the elevations extracted from the lidar DTM (with 1 m resolution) at each point's position. The geometric parameters, i.e. vehicle length/width, tree crown diameter and building length/width were directly measured on the rectified images in a Geographic Information Systems (GIS) environment. Three different classes of objects were separately investigated:

(a) Vehicles are classified into four groups depending on their length, namely: (1) passenger and family car type - smaller than 5 m; (2) vans having lengths between 5 and 7.5 m; (3) trucks with lengths between 7.5 and 10 m and (4) articulated lorries, trailers and multi-trailers usually having more than 10 m. The lengths, widths and the corresponding elevations at car's roofs, of a total number of 50 vehicles are investigated. The computed mean reference heights are 1.5, 2.5, 3.7 and 4 m for family cars, vans, trucks, and articulated lorries, respectively. Related to average object height, the associated uncertainty of the manual height measurement varies between 8% for lorries and 22% for family cars for WorldView-3 and between 34% and 68% for Pléiades.

(b) Trees are typically classified into two categories: coniferous and deciduous. Coniferous trees are cone-shaped trees, represented mainly by spruce in this case. The second category is the broad-leaved trees with leaves that fall off on a seasonal basis, mostly represented by beech and oak. This classification was needed due to the different acquisition times: one in June (leaf-on conditions) and the other one in April (leaf-off conditions). The diameter and elevations were measured for 100 trees (50 trees for each category: deciduous and conifer-

ous). Depending on crown diameters, they were grouped into seven categories, beginning with trees with a crown smaller than 2.5 m, ending at large trees with 10 m diameter. The computed mean reference heights for the seven coniferous tree classes are: 5.5, 7.8, 11.0, 14.6, 17.2, 23.4, and 28.7 m, with uncertainties between 1% and 6% from object height for WorldView-3 and between 5% and 24% for Pléiades. The mean reference heights for the deciduous tree classes are: 3.1, 5.4, 8.0, 12.6, 15.4, 16.2, and 18.5m, with uncertainties between 2% and 10% from object height for WorldView-3 and between 7% and 44% for Pléiades.

(c) For buildings, two geometrical parameters are taken into account: length and width. According to their size, built-up structures are grouped into several classes starting with very small (5 m in length and width) to large (50 m length and 25 m width). Therefore, lengths, widths and roof ridge elevations were measured for a number of 100 buildings in both Pléiades and WorldView-3 images. The mean reference height values vary from 2 m (small built-up structures) to 10-12 m (large industrial buildings), with associated uncertainties from 2% to 16% of from object height for WorldView-3 and between 11% and 68% for Pléiades.

While identical trees and buildings were investigated in both Pléiades and WorldView-3 images, this was not possible for vehicles, since they are moving objects. Therefore, (parking) vehicles were randomly selected within the entire scene also using the non-overlapping parts.

4.6.2 Object height differences

For a clear height estimation, they were considered in this work only single objects, located in free, open and smooth areas, without any other entities in their close surroundings. Assuming that in the free, open areas, the DSM coincide with DTM elevations, the vertical quality of the photogrammetric derived surface models was inspected by computing the vertical offsets from the reference lidar DTM (1 m resolution and $\sigma_z = 0.12$ m). The results showed a good correspondence of the DSMs with the reference DTM, featuring a RMSE of 0.32 m for Pléiades and of 0.20 m for WorldView-3. When it comes to individual objects, sometimes a constant offset could be observed, but it was always below 0.30 m. Therefore, it did not significantly impact the height investigations. The reconstructed individual object height was extracted from the interpolated surface models. Heights of buildings, vehicles and trees were computed by subtracting the DTM elevations from the elevations of the highest object points (located on roof ridges, on car roofs, treetops) at the same 2D location.

The reference height (H) was considered as being the real object height, which was computed by subtracting the DTM elevation from the manual measurements (Section 3.2) at the same position.

$$H = Z_M - Z_{DTM} \quad (m) \tag{4.3}$$

$$h = Z_{DSM} - Z_{DTM} \quad (m) \tag{4.4}$$

with H the reference object height, h the estimated object height, Z_M the manual elevation measurements, Z_DTM the elevation of the ground surface and Z_DSM the elevation of the reconstructed DSM. In Equations (4.3) and (4.4) the Z_DTM

elevation values are identical, since for computing the reference and the estimated heights, the same ground elevation at the same 2D position was considered. Based on the defined equations, for each individual object, were obtained a reference height (H) and two estimated heights (from Pléiades and WorldView-3 DSMs). These values will be employed in further analysis.

4.6.3 Height estimation of single objects

Within this investigation, the aim was to estimate the Pléiades and WorldView-3 DSMs heights at single objects as a function of object type and size. For this purpose, three object types were analysed: non-moving vehicles, trees and buildings. The main parameters taken into account are vehicle length/width, tree crown diameter, and building length/width. For each single object, the estimated heights were compared with the monoscopic reference measurements based on a ratio measure. The percentage value derived by the following equation describes how much of the reference height was reconstructed:

$$p(\%) = \frac{h}{H} \cdot 100 \qquad (4.5)$$

with H the reference height, h the estimated height and p the reconstruction percentage.

4.6.3.1 Vehicles

By a visual inspection of the reconstructed Pléiades-based 3D point cloud, it was observable that points corresponding to small vehicles (vehicles with lengths smaller than 5 m) were not elevated against their surroundings. Larger vehicles were reconstructed in height, but not entirely (Figure 4.53).

Specifically, Figure 4.53 represents a truck with a length of ∼10 m and a real height of 3.5 m that is visible in both Pléiades (a) and WorldView-3 images (b). On the left-hand side are the magnified satellite image details of the vehicle and on the right-hand side are the corresponding reconstructed height color-coded DSM. From the DSM and the cross profiles (c), it is clearly visible that in contrast to Pléiades, where there is only ∼1m elevation change, the WV3-based profile reconstructs the vehicle height up to 2.9 m, which represents ∼83% of the real height.

Using Equation (4.5), the height percentages of 50 measured vehicles were computed for both cases: Pléiades and WorldView-3 sensors. Vehicles were classified into four different groups depending on their lengths, starting with small family cars, of approximately 5 m, followed by vans, trucks, and finally by large articulated lorries with more than 10 m length. For each group, containing between 10 and 13 vehicles, the mean value of the reconstruction height percentage and the standard deviation were determined. They are shown in Figure 4.54a for both Pléiades (blue) and WorldView-3 (yellow) data. By increasing length, the height percentage also increases, reaching up to 60% (Pléiades) and 92% (WorldView-3). In the case of the Pléiades imagery, small vehicles (<7 pixels length in image) do not show any height information while for family cars, vans, and trucks the reconstruction height percentage is less than 30%. Only for large vehicles (lengths >10 m, 14 pixels in image) it exceeds 50%. On the other hand, in WorldView-3 DSMs vehicles have a good height reconstruction (reaching over 90% for lengths >10 m, 33 pixels in
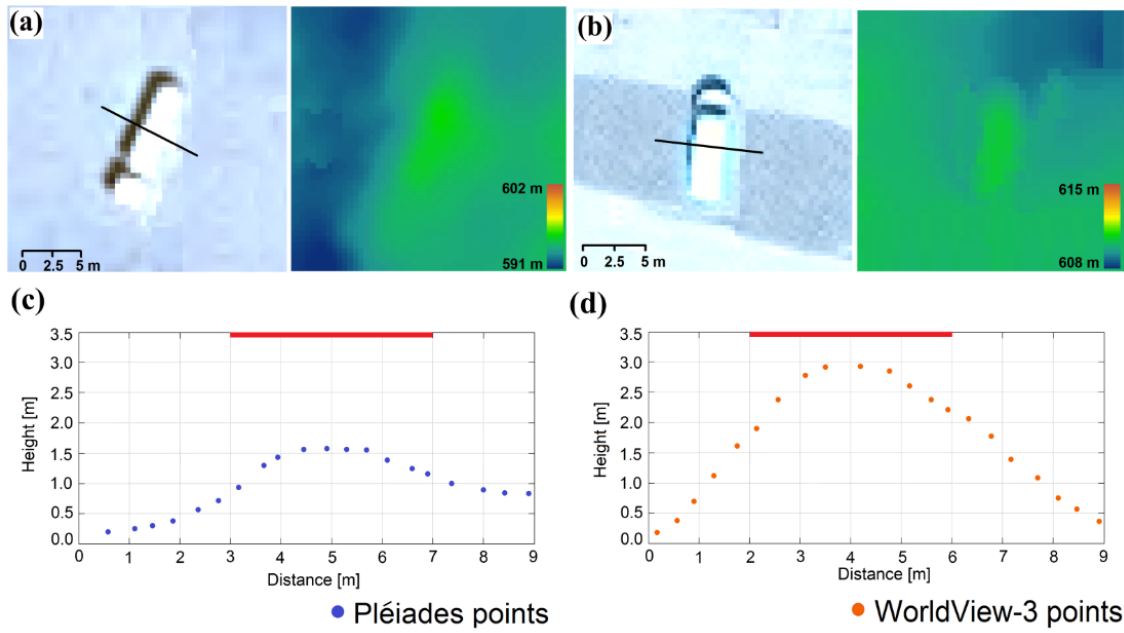
Figure 4.53: Height estimation for a truck. (a) Pléiades satellite image detail with profile location (left) and reconstructed DSM Z color-coded (right); (b) WorldView-3 satellite image detail with profile location (left) and reconstructed DSM Z color-coded (right); (c) Pléiades- and (d) WV3-based vehicle profiles of 0.5 m width each from the matched point cloud; the red horizontal line is the corresponding measured reference height.

image). An exception are family cars ($\sim$16 pixels length), which have a percentage of less than 40%. The standard deviations of the estimated heights are notably smaller than the overall trend and they become smaller with increasing vehicle length.
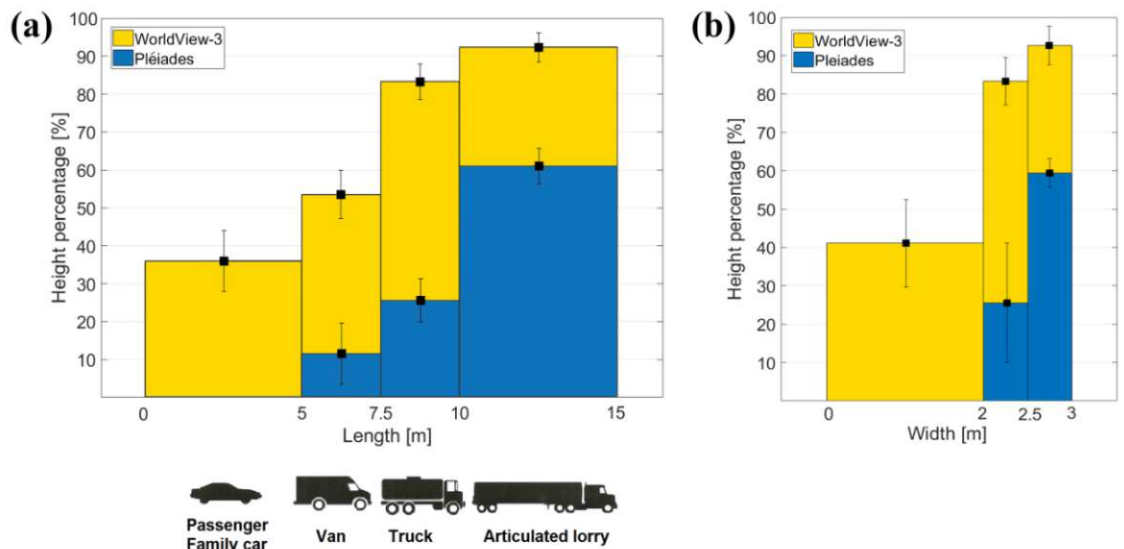

Figure 4.54: Interval-based percentage and standard deviation for height estimation of vehicles depending on (a) lengths and (b) widths.

A similar analysis was conducted by using vehicle width as main parameter for

117

vehicle grouping (Figure 4.54b). Because small family cars and vans have similar widths (∼2 m), they form a single class, followed by trucks (2 to 2.5 m widths) and articulated lorries (2.5 to 3 m widths). Based on the WorldView-3 DSMs vehicles with up to 2 m widths get ∼41% of their real height reconstructed. This value is higher when compared with the percentage of the first length-class (37%) because it contains both family cars and vans, leading to an increased mean height percentage and standard deviation, as well. The mean height percentages reach 59% (Pléiades) and 92% (WorldView-3) for the very large vehicles. The increased values of standard deviations suggest a higher variation of the reconstruction height percentages within the considered width-classes. Therefore, the length parameter is more suitable for the description of vehicle's height estimation.

4.6.3.2   Trees

Trees were investigated in two separate classes: deciduous trees, located especially on roadsides, parks and cities, and coniferous trees, found in forested areas and forest patches. Only single, isolated trees, not close to buildings or any other objects were taken into consideration. In this case, the visible crown diameter was used as geometric parameter. One hundred trees were investigated and the same formula for height percentage computation, Equation (4.5), was applied. Based on diameter, they were grouped into seven different categories, each containing approximately 12 trees. The mean reconstruction percentages gradually increase with crown diameter and the standard deviations decrease (Figure 4.55). For WorldView-3, the heights of coniferous trees with crown diameters <2.5 m (8 pixels in image space) are reconstructed less than 50%, whereas trees with crown diameters larger than 5 m (16 pixels in image space) are over 95%. Generally, heights of coniferous trees are better estimated from WorldView-3 than from the Pléiades DSMs, where trees with diameters >7.5 m (11 pixels) barely reach 75% height reconstruction. For deciduous trees (Figure 4.55b) it can be observed that the situation is reversed: they get better height information from Pléiades DSM, than from WorldView-3 DSM. This is mainly because of the different acquisition times: June (Pléiades) and April (WorldView-3), indicating leaf-on and leaf-off conditions which clearly influence the appearance and texture of the trees in the images. An example can be seen in Figure 4.57a, where only tree stem and branches are visible in the WorldView-3 image resulting in a poor dense image matching quality. Only for trees with crown diameter lager than 9 m, heights just over 50% of the true value get reconstructed.

When analysing the two tree types only from the Pléiades images (Figure 4.55), it could be seen that there is a slightly better height reconstruction for the very small and large deciduous trees, compared with the coniferous. The small deciduous trees in the first group (with crown diameters smaller than 2.5 m) have 26% of their real height reconstructed, whereas the small coniferous trees in the first group have only 21%. The same is for the last group, where for the large deciduous trees (with crown diameters bigger than 8.75 m), the percentages are very close to 80%, whereas for coniferous they reach only 75%. On the other way, for the coniferous trees of medium sizes (with crown diameters between 2.5 and 7.5 m) they was obtained better height reconstruction percentages than for the deciduous.

In the following, two examples of height estimation of a coniferous (Figure 4.56) and a deciduous tree (Figure 4.57) are shown. They are comparatively analysed for
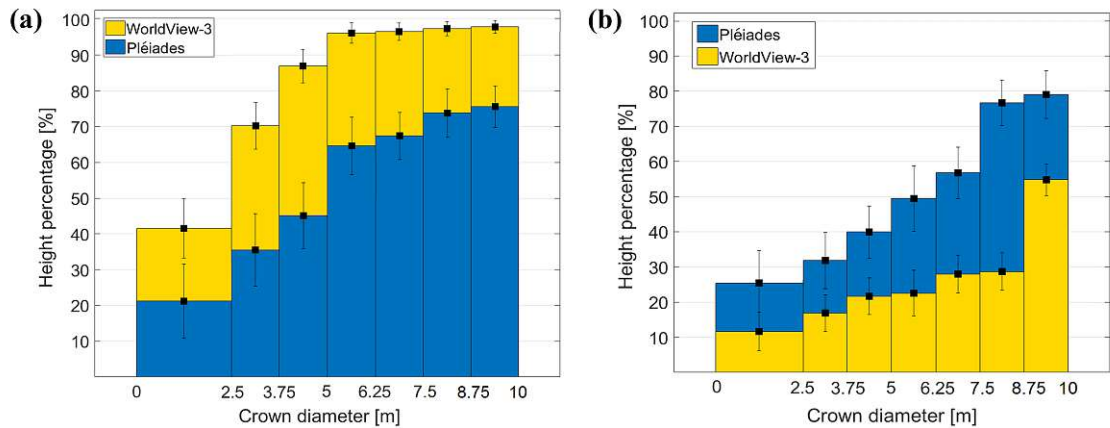
Figure 4.55: Interval-based percentage and standard deviation for height estimation of single trees (a) coniferous trees and (b) deciduous trees.

Pléiades and WorldView-3 satellite images based on both DSM and point cloud. For single coniferous trees with a 7 m crown diameter (Figure 4.56) different results were achieved, based on the input data. From the true height of 12.2 m, the estimation resulted in approximately 12 m for WorldView-3 and only 8 m for Pléiades, which represents 98% and 65% of the real height.
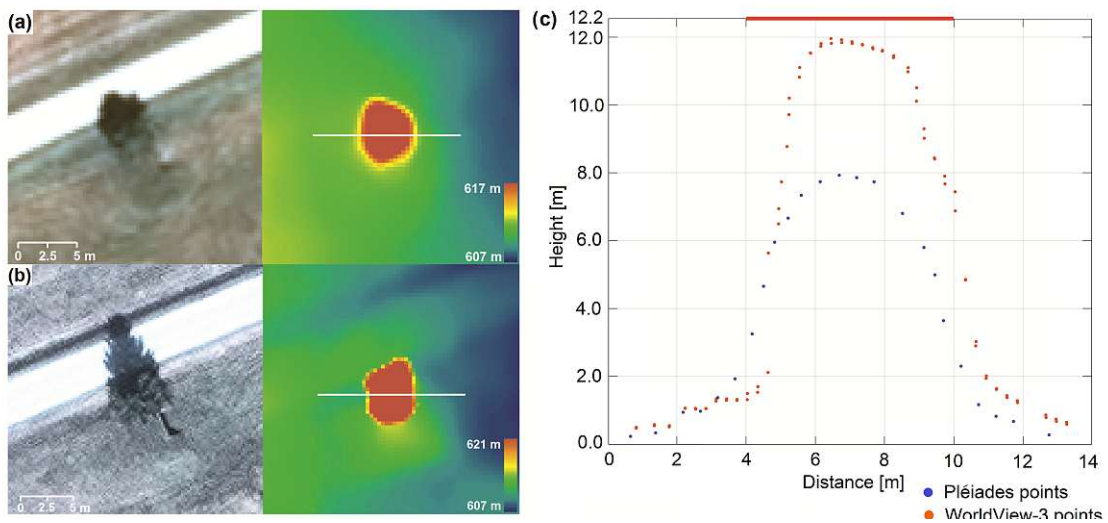

Figure 4.56: Height estimation of a coniferous tree with 7 m crown diameter: (a) Pléiades satellite image detail (left), corresponding height-coloured DSM and profile location (right); (b) WorldView-3 satellite image detail (left), height-coloured DSM with profile location (right) and (c) Tree profile of 0.5 m width from the matched point clouds; the red horizontal line is the measured reference height.

A small area of the DSMs and the 3D point clouds generated from the Pléiades and WorldView-3 satellite imagery are shown in Figure 4.57. The largest deciduous tree with leaves (Figure 4.57a first row) has a reasonable appearance in the DSM and an estimated height of 14 m. Meanwhile, the same leafless tree in the WorldView-3 image (Figure 4.57a second row) has a height of only approximately 4 m. A visual analysis of height differences for this deciduous tree and a small building (next to

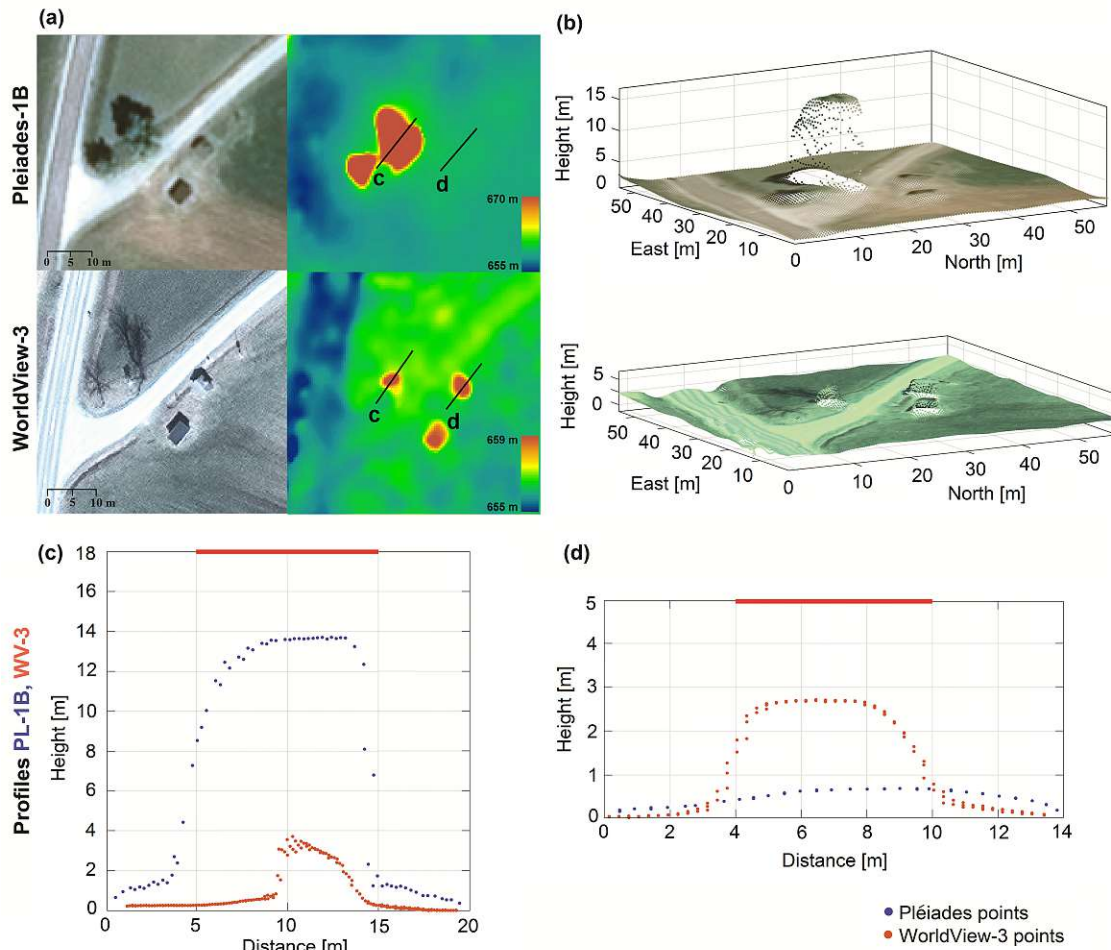the tree) is shown in two profiles (c) and (d) in Figure 4.57.

Figure 4.57:   Height estimation of a deciduous tree and a small building: (a) Pléiades (first row) and WorldView-3 (second row) views with corresponding height - colour coded DSMs; (b) RGB coloured 3D point clouds, 1 m width height profiles from the matched point clouds; (c) for a deciduous tree; (d) for a small building; the red horizontal lines are the corresponding measured reference heights.

4.6.3.3   Buildings

The third category of investigated objects were single buildings, usually met in suburban and rural areas. They usually had regular geometric shapes, such as rectangles or squares, which facilitated the visual identification of the corresponding dimensions in the rectified images. The building length / width were measured in the rectified images, based on their roof top view. For buildings with a T- or L-shaped plan, the length was considered as the longest dimension between extreme roof edges, whereas the width as the distance across the roof from one side to the other. One hundred buildings were chosen to cover a wide range of dimensions from 1 to 50 m length (large industrial buildings) and widths from 1 to 25 m. In the example shown in Figure 4.57b, the two small buildings are elevated against their surrounding in the WorldView-3 data, which is not the case for the Pléiades data. In addition, profile (d) with two overlaid point clouds corresponding to a 3.6 by

2.5 m building clearly reveals the higher potential of WorldView-3 sensor for height reconstruction, as opposed to Pléiades. Even if the building edges appear very smooth, the corresponding main height is still reconstructed. Again, Equation (4.5) was applied for investigating the potential of Pléiades and WorldView-3 DSMs for building's height estimation with respect to their lengths and widths (Figure 4.58). Based on their length, they were defined ten different categories, each containing approximately ten buildings (Figure 4.58a). Like in the previous analyses for vehicles and trees, the computed mean height percentages for buildings show a similar trend, values gradually increasing with building length. In case of the Pléiades imagery, buildings with lengths smaller than 10 m (14 pixels in image space) have a height reconstruction percentage of less than 30%, whereas buildings with lengths greater than 5 m have percentages beyond 50%. For both, Pléiades and WorldView-3, buildings with lengths >43 pixels get over 90% height reconstruction.
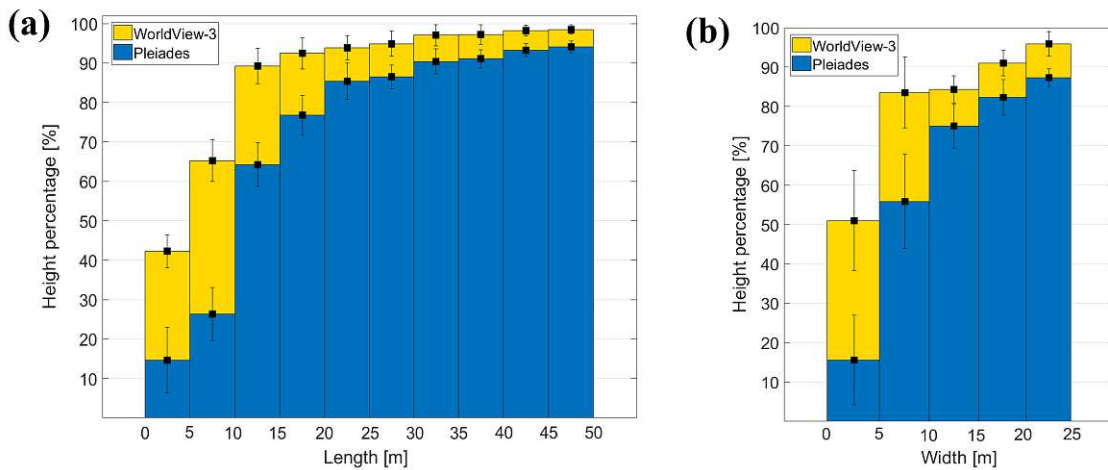

Figure 4.58: Interval-based percentage and standard deviation for height estimation of buildings depending on (a) lengths; (b) widths.

Due to the lower variation in building-width, only five intervals, starting from 0–5 m to 20–25 m were considered (Figure 4.58b). Compared with height-percentage analysis based on building length, for the width parameter the situation was different. When buildings were grouped based on their corresponding widths, then an increased number of items within each interval (only 5 width-intervals) is obtained. This lead to a higher variation of height estimation percentages and hence to higher values for the standard deviations within each interval. When analysing the reconstruction height percentages based on the WorldView-3 images for both 0–5 m and 5–10 m length/width intervals, higher mean values for the reconstruction percentages computed based on buildings widths were observed. Nevertheless, the corresponding standard deviations are larger in contrast to those computed on the length basis. The standard deviations for the length-intervals (Figure 4.58a) become smaller with increasing building length, showing a clearer trend, therefore the length parameter gives a better estimation when analysing the height reconstruction of buildings.

The two scatter plots (Figure 4.59) show the relationships between building length and width and the reconstruction height percentage. As expected, height accuracy is better for WorldView-3 images than for the Pléiades data. The delineation lines
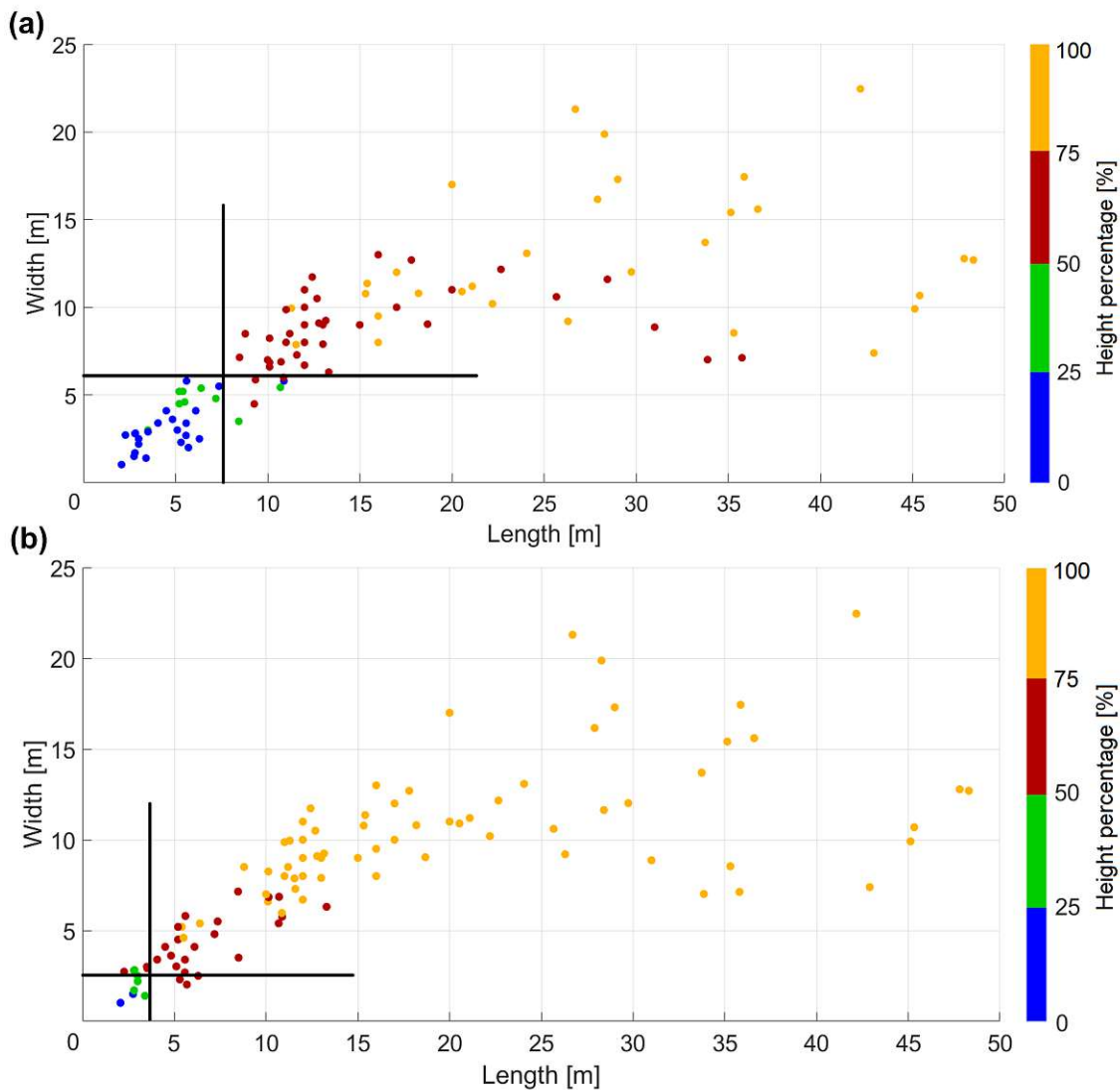
(a)

(b)

Figure 4.59:   Height estimation of buildings: (a) for Pléiades; (b) for WorldView-3.

separate the buildings with higher and lower reconstruction height percentage (p) than 50%. If for Pléiades data a 2D minimum building size of 7.5 m length by 6 m width is needed in order to get p > 50%, for WorldView-3 data only 4 m x 2.5 m is needed. The mean value for the residual heights corresponding to reconstructed buildings with p > 50% (of 2.02 m for Pléiades with 0.7 m GSD) is comparable with the residuals of 1.94 m obtained by Moreira et al. 2013 when using a DSM from a Geo-Eye stereo pair (0.5 m GSD).

4.6.4 Capability for small isolated object height estimation

In order to investigate the potential of Pléiades and WorldView-3 tri-stereo DSMs for single object's height estimation 50 vehicles, 100 trees and 100 buildings were analysed. Vehicle length/width, tree crown diameter and building length/width are identified as main geometric parameters for classifying single objects into different groups. From the performed analyses and investigations, it clearly results that acquisition geometry has an effect on accuracy, but the object height estimation based

on automatically derived DSMs primarily depends on object size. Results show a gradual increase of the estimated heights with object dimensions. Starting from very small objects (whose heights are not reconstructed) the values for the height reconstruction percentage reach almost 100% for very large single objects (e.g. buildings with lengths >35m in the case of WorldView-3 DSM). Hence, individual object's reconstructed heights strongly depend on object type, size and the sensor being used. For small vehicles, buildings and trees (lengths < 7 pixels, crown diameters < 4 pixels) the Pléiades derived DSM showed no or less than 20% of the actual object's height. For the large vehicles, buildings and trees (lengths > 14 pixels, crown diameters > 7 pixels) the estimated heights were higher than 60% of the real values. In the case of the WorldView-3 derived DSM, the estimated height of small vehicles, buildings and trees (lengths < 16 pixels, crown diameters < 8 pixels) was less than 50% of their actual height, whereas larger objects (lengths > 33 pixels, crown diameters > 16 pixels) got reconstructed more than 90% in height. If a percentage threshold p = 50% for an appropriate height reconstruction is considered, then the following minimum size values for the single objects are obtained: 10 m (Pléiades) and 5 m (WorldView-3) for vehicle lengths; 5 m (Pléiades) and 2.5 m (WorldView-3) for tree crown diameters and 7.5 x 6 m (Pléiades) and 3.5 x 2.5 m (WorldView-3) for building length/width. The ratio of these dimensions fits well to the GSD ratio of 70 cm to 31 cm.

Generally, for all single objects except deciduous trees, the results achieved with WorldView-3 data are better than those achieved by using the Pléiades data. Hence, the ability of WorldView-3 derived DSM for height estimation of single objects is higher compared with Pléiades DSM. This is mainly because of the very high resolution (0.31 m) that leads (through photogrammetric processing) to very dense 3D point clouds, describing the real ground surface and the objects on it more accurately. The acquisition times also play a significant role in the photogrammetric procedure. For stable objects such as buildings, the different periods of acquisition do not affect the 3D reconstruction, but for deciduous trees, the automatic reconstructed heights are highly underestimated in the leaf-off season. Their leafless appearance in the images brings difficulties to the image matching process when finding correct correspondences for tree branches. From this observation, it can be concluded that the image appearance and texture of trees (and other single objects) are also very important for the 3D reconstruction and height estimation. For all investigated objects, the resulting histograms (Figure 4.54, Figure 4.55, Figure 4.58) show interesting trends, opening new investigation options for refinement such as adding the influence of object orientation, colour and texture in the feature reconstruction ability of heights. The poor performance on small individual objects is mainly caused by the continuity constraint of dense matching algorithms.

The estimation of single object's heights depends on the accuracy of satellite image orientation, but also on the elevation accuracy of the derived DSMs and on the used lidar reference DTM. The vertical accuracy of the reference DTM with respect to the RTK GCPs is 0.12 m and the accuracy of the reconstructed photogrammetric models with respect to the DTM are 0.32 m for Pléiades and 0.20 m for WorldView-3 in the open, smooth areas surrounding the analyzed objects. The vertical accuracy of the DSMs from VHR satellite imagery is influenced by the acquisition geometry. The Pléiades scenes having a narrow convergence angle of 12° show a lower vertical performance compared to WorldView-3 scenes, which have a 23° convergence an-

gle on the ground. In addition, the smoothness effect of the robust moving planes interpolation for deriving the DSMs is at sub-decimetre level. Under these conditions, the resulting height accuracy is reasonable and does not significantly influence the investigations, since the measured single objects are usually higher than 1.5 m. Based on the different GSDs and B/H ratios of the two satellite systems, it can be said that σ_o of block adjustment, RMSE of check point elevation error and height reconstruction success for individual objects are similar, with deviations of up to 20%.

4.7 Visibility analysis from satellite imagery

4.7.1 Street visibility in Barcelona WorldView-3 images

Because the satellite position along its orbit has different in-track and cross-track viewing angles, there are areas in the images occluded by high buildings, trees, and other artificial objects on the ground.

Being the second most populous municipality of Spain, Barcelona is a large city with tall buildings, but currently not exceeding 150 m in height. From a thorough analysis of the derived DSM, it follows that typical buildings in the city centre have 30 m height. Moreover, building have similar rectangular atrium-shapes following the regular distribution of street directions. This was double checked with additional visualisation of the satellite images and OSM data. In the city center, the street-network has a conventional grid pattern, with arterial roads and local streets parallel and orthogonal to each other, forming a pattern of squares (Figure 4.60).
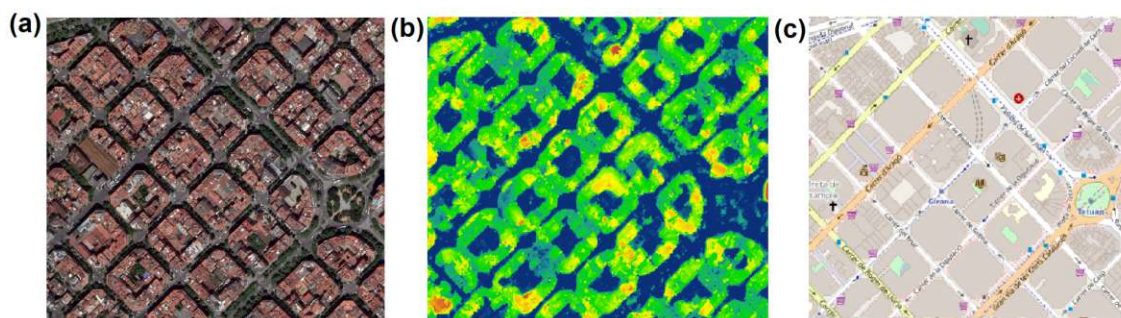

Figure 4.60: Barcelona street network type (grid pattern with streets that intersect at right angles). Detail view of: (a) WV-3 forward image, (b) DSM 0.30 m resolution with filled occlusions by interpolation, (c) OSM street view.

4.7.2 Satellite viewing geometry

The visibility potential in satellite images was investigated in relation with satellite viewing geometry. The most important acquisition angles are depicted in Figure 4.61 and are defined as follows:

Incidence angle ν_C (global incidence) is the angle between the normal on the ground and looking direction from the satellite.

Viewing angle β is the angle between the looking direction and Nadir direction of the satellite.

Azimuth angle α_T is the angle between geographic north and the and satellite

track on the ground. The range for the azimuth angle is [0, 360 degrees], measured clockwise positive.

While the viewing angle is defined in space, both azimuth and incidence angle are calculated on the ground (Astrium 2012).
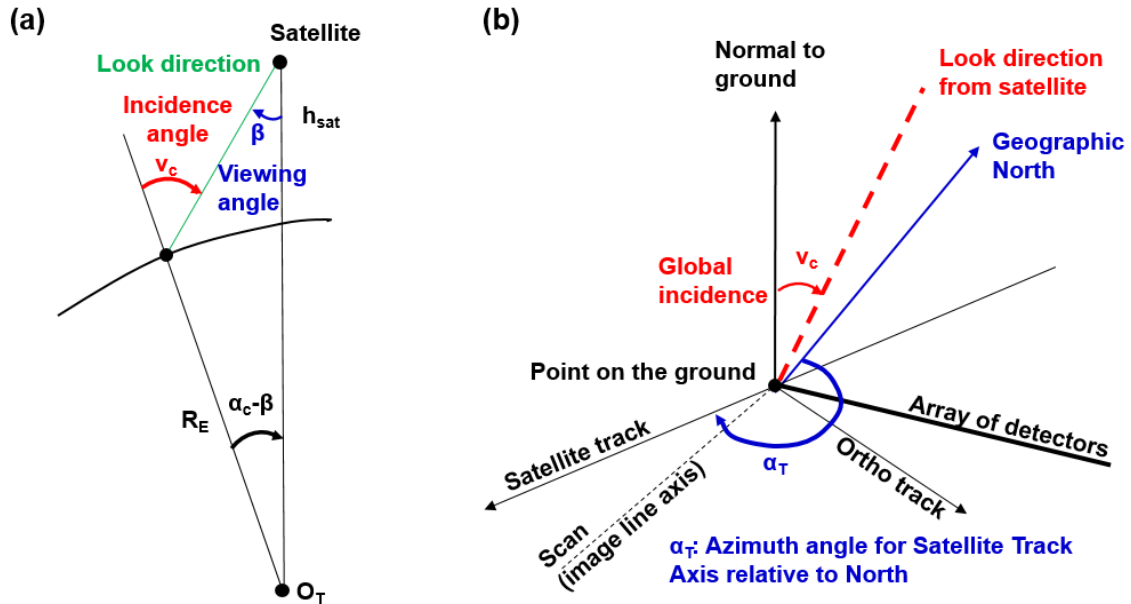

Figure 4.61: Satellite acquisition angles: (a) relation between incidence and viewing angle; (b) incidence angle and azimuth of the satellite track (adapted from Astrium 2012).

The visibility analysis was performed on the WorldView-3 satellite imagery in Barcelona urban area, where the typical distance between two buildings in the city centre is approximately 20 m. This covers not only the street width (10 m), but also the sidewalks and parking areas on both street sides. Figure 4.62(a) illustrates a profile-view for visible and non-visible areas during image acquisition, where a particular case is being considered, with the satellite track parallel with the street azimuth and an in-track viewing angle of 0°.

The street visibility can be analysed with the incidence angles (ν_c), azimuth of the camera sensor (α_c) and street direction (α_s). According to Figure 4.62(b), the non-visible street areas ω_o can be described by both, geometry of the city (h_b, α_s) and satellite viewing geometry (α_c, ν_c), with the following formula:

$$\omega_o = h_b \sin(-\alpha_s + \alpha_c) \tan \nu_c \qquad (4.6)$$

Here ω_o is the non-visible distance on the ground hidden by buildings, h_b the building height, α_s and α_c the street and camera azimuths, and ν_c the incidence angle.

The incidence angle (ν_c) is defined as the angle between the ground normal and looking direction from the satellite sensor (note that those angles are defined at the ground, whereas the Nadir angle at the satellite may be different by a few degree because of Earth curvature). For the two Barcelona satellite images the incidence angles are 20° and 24°. The azimuth of the camera sensor (α_c) is the angle between North and the viewing direction projected on the ground, clockwise positive from
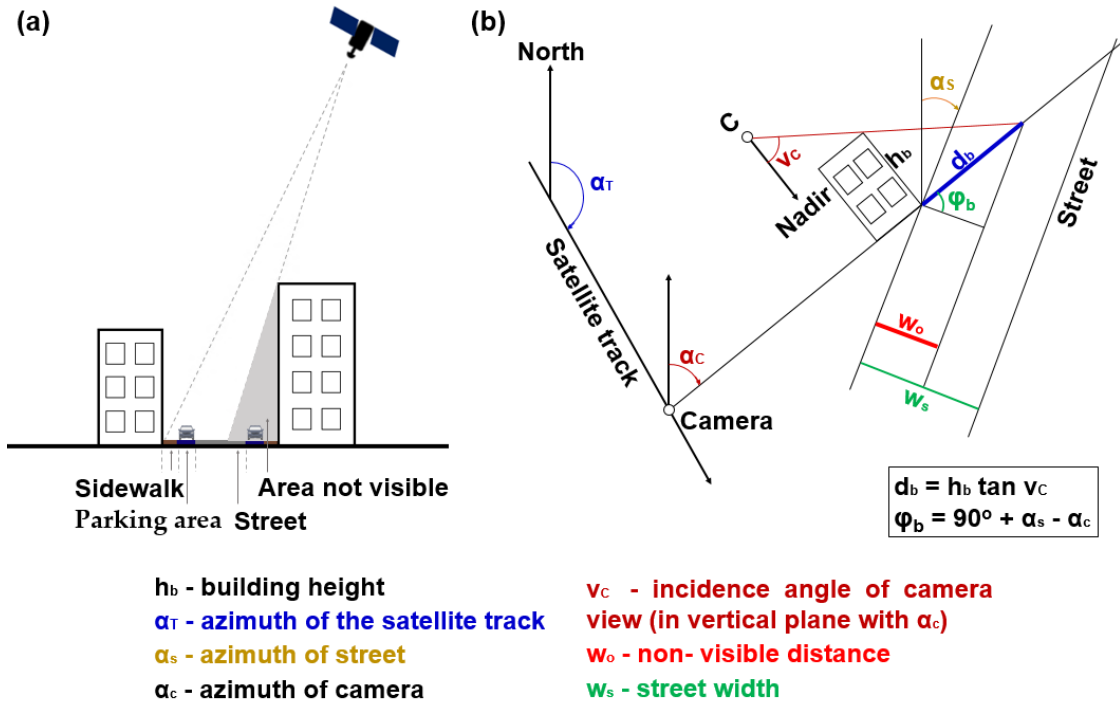
(a)

(b)

Figure 4.62: Satellite image acquisition: (a) profile-view illustration of visible and non-visible areas (satellite track parallel with street azimuth and a null in-track viewing angle); (b) satellite viewing geometry for street visibility.

0° to 360°. The Barcelona stereo images have azimuth angles of 228° and 345° for the forward and backward scenes, respectively.

Streets located in the central area of the city have a regular distribution with right angle intersections (Figure 4.62). Hence there are mainly two types of streets: (a) with a south-west to north-east direction ($\alpha_s = 45°$) and (b) with a south-east to north-west direction ($\alpha_s = 135°$). The building height h_b is 30 m.

For computing the visible part of the streets, the street width (ω_s) needs to be considered, too. The visibility can then be computed by subtracting the non-visible distance (ω_o) from the street width (ω_s):

$$\omega = \omega_s - \omega_o \tag{4.7}$$

By applying the above equations and considering a street cross-section, different values for the two types of streets in the two images are obtained. In the forward scene, the visible distances are 19.4 m for streets with 45° azimuth and 9.1 m for streets with 135° azimuth, i.e. visibility of 97% and 46% for north-east and north-west oriented streets, respectively. For the backward image, the resulting visible distances are 8.3 m and 13.2 m (visibility 41% and 66% for the north-east and north-west oriented streets, respectively). Thus, at least the western halves of the streets are visible in both images.

Obviously, the smaller the incidence angles, the better is the street visibility. Therefore, to avoid reduced visibility and to have as few occlusions as possible in the images, the acquisition collection parameters should be taken into account when ordering satellite imagery in urban areas. Concerning occlusions caused by buildings, small in-track and cross-track angles are recommended for a better visibility. The

acquisition geometry can be specified in the technical documentation when ordering satellite images (additional costs are charged). In case of archive images, there is typically hardly any choice of viewing directions. Using two image pairs, one looking eastwards and one looking westwards, would allow observing the entire parking space with short interval in-track imagery. To reduce the occlusions caused by trees, images can be acquired during leaf-less periods, however the low sun elevation angle leads to a lower image quality under these circumstances.

5

Summary and Conclusions

Satellite and airborne photogrammetry together with lidar are proven techniques for 3D geospatial information acquisition of the Earth's surface. Individually, each technique has its own advantages, but also drawbacks. With Airborne laser scanning very precise and accurate 3D point clouds can be obtained, with the main benefit of vegetation penetration making possible the underlying ground modelling and a second main benefit is that no image matching required, since the 3D information is directly collected, therefore having a faster acquisition. However, costly navigation sensors (GNSS/INS) and time-consuming post-processing trajectory georeferencing and registration procedures are needed. On the other hand, aerial images have the advantage to record large surface areas from a single position in time, with spectral information in at least three channels. Through photogrammetric procedures, i.e. dense image matching, very dense point clouds are derived.

In contrast to the aforementioned acquisition methods, the stereo/tri-stereo satellite image acquisitions have the great advantage and potential of covering very large surface areas in a timely and cost-effective manner from very high positions in space. The agile very high resolution satellite sensors such as Pléiades and WorldView used in this thesis (with ground resolutions of 0.7 m and 0.3 m for the panchromatic band) are able to capture areas of interest from different viewing angles within a single flight. This is the fundamental requirement for the three-dimensional scene reconstruction. Although satellite systems have significantly lower ground resolution than airborne or drone-borne sensors, satellites may capture any point on earth at regular time intervals. Additionally, due to their high temporal resolution and coverage of inaccessible or difficult to reach areas (for example due to natural disasters, remote locations or crisis) many new applications that would not be achievable with other measuring techniques, become possible with satellite imagery. Moreover, the very high spatial resolution of the newer optical satellite sensors (WorldView-3, Pléiades Neo with 0.3 m) makes satellite images comparable to airborne ones. When thinking at their location in the orbit, the high altitude of the satellites (694 km and 617 km for Pléiades and WorldView-3, respectively) impacts the reconstructed point clouds, which have an appearance more of a 2.5D than a 3D space. The DSMs from satellite imagery mostly describe the upper surface of the ground with geometry of objects on top of it. However, in contrast to the classic two-dimensional approach, which consists only of satellite image interpretation, the ability of 3D information extraction from the stereo/tri-stereo image acquisitions is clearly an important advantage.

In this dissertation the potential and limitations for 3D information extraction and point cloud classification from VHR optical stereo/tri-stereo satellite imagery of Pléiades and WorldView-3 are investigated (Sections 4.2.1 and 4.5). To be able to create an overall big picture for the satellite image capabilities, four distinct areas

of interest with particular topographic characteristics and located in different countries (Austria, Spain, Sweden and Nepal) were selected: Allentsteig (land landscape with open agricultural fields and grasslands, forests and rural areas), Barcelona and Stockholm (two dense urban areas with high buildings, infrastructure, industrial zones, green spaces and port areas), and Himalaya (the world's highest and most spectacular mountain range with peaks, valleys, steep slopes, snow and glaciers). Their unique features brought specific challenges to the photogrammetric processing chain. The first in-situ test area, Allentsteig, located in northern Lower Austria, was subject to a more detailed accuracy analysis, georeferencing improvement and satellite geometry image correction, since here precise reference data were available: laser scanning terrain model, orthophoto from aerial images and locally measured ground control points.

The entire photogrammetric workflow, comprising satellite image triangulation with RPC bias correction based on GCPs (where available), dense matching and 3D reconstruction with detailed investigations and results concerning the quality and accuracy assessment are described in Sections 4.1 and 4.2. The resulted photogrammetric satellite-based point clouds in open areas with fields and grasslands have a good geometry, with regular distributed points. However, smoothing effects appear at buildings roofs edges that look bevelled, due to the rapid elevation change. Urban environments are most problematic, due to their complex morphology, image acquisition geometry and urban texture. Here, the reconstructed point clouds show elevation discrepancies (up to several meters) especially at narrow roads and urban canyons, but also areas with missing elevation information (occlusions), due to high buildings where large elevation differences to the surrounding ground occur. However, for applications where the overall elevation of urban features is more important than the precise reconstruction of the shape, the quality of the reconstructed point clouds from WorldView-3 images can be considered optimal. Nevertheless, in urban areas, the accuracy is still not competitive with the one achievable with other techniques, such as LiDAR, especially when considering the capability of reconstructing vertical surfaces. On the other hand, in the Himalaya mountains, dense image matching encounters difficulties in areas with low image texture (white snow, ice), where miss-matches appear as outliers, noise and scattering effects.

The resulted three-dimensional point clouds serve as database-inputs for further products such as digital surface models and orthophotos. The absolute vertical accuracy of high resolution DSMs was evaluated against ground checkpoints, lidar-derived DTM (where available) or ALOS DSM. The focus is on the assessment analyses in free, open and smooth areas, without any vegetation or artificial structures like buildings or infrastructure and for this, a mask derived from available lidar DSM and DTM was used. Generally, when GCPs are employed in satellite image triangulation, then the vertical accuracy of satellite-based DSMs is less than 1.5 pixels, whereas it decreases to more than 25 pixels when no RPC-bias compensation with GCPs is performed. In the latest case, the satellite-driven DSMs show systematic errors with elevation offsets and RMS values up to 27 m.

To study the impact of different acquisition geometries, image pairwise DSMs (FNB, FB, FN, NB) were generated for both Pléiades and WorldView-3 datasets in Allentsteig. The vertical quality of the reconstructed DEMs was analyzed with traditional and robust accuracy measurements, resulting in non-Gaussian distributions of errors, with a RMSE of 0.96 m (1.4 pixels) for Pléiades and of 0.37 m (1.2 pixels)

for WorldView-3. When compared to ground truth lidar DTM, the elevation differences show an undulation (1.5 pixel), similar to waves that are visible in the along track direction. In order to minimize this systematic effect and the vertical error caused by horizontal and vertical offsets, the photogrammetrically derived DEMs are aligned to the reference DTM by applying an affine 3D transformation determined with the LSM techniques. The results show improvements in the vertical accuracy to 0.61 m (0.9 pixels) and 0.24 m (0.7 pixels) for Pléiades and WorldView-3 tristereo scenes, respectively, however the "wave-effect" is still present, with a slight amplitude decrease to approximately one pixel.

In this thesis it is proposed an approach for the analysis and estimation of VHR satellite image distortions caused by attitude oscillations in flight direction, by the use of a stereo triplet with vendor-provided RPCs and a high resolution elevation model. For this purpose, three Pléiades images were tested and the experimental results demonstrated the effectiveness of the proposed method, which improves the satellite image geopositioning accuracy. Additionally, a thorough analysis regarding satellite-based DSMs accuracy analysis and different improvement strategies of the Pléiades DSMs were conducted (Sections 4.2 and 4.3). From the performed experiments, it can be concluded that in flying direction, the geometric accuracy of Pléiades images depends on the sensor attitude, which is apparently affected by satellite oscillations. When compared to a Lidar high resolution elevation model, the computed satellite-based DSMs show periodic systematic height errors as undulations (similar to waves with a maximum amplitude of 1.5 pixels) visible in along-track direction. This suggests that image orientations are not sufficiently determined by employing a common number of GCPs for RPC bias-correction. The periodic vertical offsets in the computed DSMs could not be effectively compensated even if the number of GCPs was increased to 300. This strategy brought improvements to the vertical accuracy of the Pléiades DSMs with 20% in the overall RMSE, which implies that the accuracy in height is sensitive to the number and distribution of GCPs. Nevertheless, the systematic elevation offsets are preserved. Similar to the 300 GCPs strategy, the applied global LSM technique in object space brought significant improvements to the photogrammetrically derived DSMs, since the RMSE were reduced by 26%. However, the systematics in-track direction were still present. The preservation of systematic height errors in the computed satellite-based elevation models suggest a not sufficient bias-correction model for the RPCs. This is explained by the fact that in flying direction, satellite image geometry highly depends on the accuracy of the sensor orientation angle. Hence, a quick change in viewing direction leads to sensor vibrations, which can not be captured by the bias-compensated 3rd-order rational polynomial coefficients. The proposed approach based on corrections in image space can detect and estimate the periodic image distortions in-track direction. With amplitudes of less than 0.10 pixels, oscillation period (T) of 0.70 seconds, and frequency of 1.42 Hz, the image corrections describe actually the small vibrations of the Pléiades satellite during image acquisition with a pitch angle of $1.7 \cdot 10^{-5}$ degrees. The effectiveness of the proposed method is proven by the successfull removal of the systematic elevations discrepancies in the DSM and by the improvement of the overall accuracy with 33% in RMSE.

Overall, besides DSM quality improvement, the method presents an image post-processing scheme to eliminate the negative influence of the attitude oscillation for geo-positioning. Nevertheless, the proposed approach has two important require-

ments: a very high accurate reference model and a landscape which is (at least) partly open. It will not work for areas with complete forest coverage.

The improved satellite-based DSMs can be further used to compute object heights (Loghin, Otepka-Schremmer, and Pfeifer 2020) given a high resolution DTM, but also in many forestry applications, for example to detect tree growth (Piermattei, Marty, Karel, et al. 2018; Piermattei, Marty, Ginzler, et al. 2019). In this investigation, only Pléiades tri-stereo images were tested. More Pléiades images covering different areas, with distinct topography are required to further analyze the influence of satellite attitude oscillations on the accuracy of satellite-based DSMs. Additionally, further studies on other VHR satellite data sets are also necessary to evaluate the performance of the proposed approach. There are expected variations from image to image and only the character and size of deformations to be typical for a specific optical satellite, but this also depends upon the operating conditions due to high speed and fast satellite rotation.

To check the possibility of using a freely available elevation model with coarser resolution instead of the high-resolution lidar DTM, the accuracy of ALOS global dataset with 1" resolution (30 m) was investigated in Section 4.3.4. However, the results revealed systematic errors of ALOS DSM against lidar with a RMSE of 2.79 m and standard deviation of 2.32 m. Not even the LSM affine 3D transformation to lidar has brought significant improvement (a decrease of RMSE to 2.62 m).

Investigations related to satellite image orthorectification improvements are described in Section 4.4, where the focus falls on the Barcelona urban study area. Here, the true orthophoto showed distortions effects and artefacts especially along roof boundaries, caused by the high roughness of the employed satellite-based DSM. Moreover, the great altimetric discontinuities due to high buildings created the so called "double projections" or "ghost images". To improve the quality of the orthophotos, different strategies have been proposed, by replacing the DSM with either the corresponding DTM or a constant height plane. Although with a good appearance in images, the newly obtained orthophotos have the disadvantages of not handling sharp buildings edges, they are characterized by relief displacements offsets and by leaning-buildings effects. Evaluated at GCPs, the best planimetric accuracy is achieved for the true orthophoto (σ of 0.22 m), while the constant height plane-based othophoto has a displacement σ of 10.49 m. Resulted statistics for the DTM-based orthophoto shows similar accuracy to the DSM-based orthophoto, a fact explained by the location of the GCPs on the ground surface, since the terrain level is being correctly reprojected.

In this work, it was analysed the performance of a supervised machine learning classification algorithm that exploits the full 3D content of point clouds derived from dense image matching of tri-stereo VHR satellite imagery (Section 4.5). The tree based classification method has been already successfully applied to aerial laser scanning data and point clouds from aerial image matching, but in contrast to previous research, in this study, the decision tree is trained on the geometric and color features of the satellite image driven 3D point clouds. This was a challenging investigation, since the geometric information precision of the satellite-driven point clouds is lower compared to the accurate geometric 3D position of laser scanning point datasets. To evaluate the performance and quality of the classifier, a comparative analysis of the statistic metrics for the same study area acquired from three different data sources with corresponding distinct acquisition geometries was investigated (Allentsteig ac-

quired with Pléiades - 6 June 2017, Pléiades - 20 June 2017, and WorldView-3 -
8 April 2018), resulting in variations of less than 3.5% for the accuracy, precision,
recall, and F1-score. Compared to Pléiades, the higher GSD of the WorldView-3
satellite images leads to approximately three times higher processing time for fea-
ture computation, training and applying the classification model. The cross-time
performance was tested by training the classifier on a dataset and applying it on
a new dataset for the same area, but from a different acquisition time, showing a
decrease of the F1-score below 3%. The low variations confirm not only the stability
of the adopted classifier, but also the transferability performance and stability of
the sensor in terms of spectral geometric precision and geolocation accuracy.

In addition to general geometric accuracy investigations and classification of the
reconstructed data, special attention was paid to small individual objects that are
at the limit of satellite image ground resolution. Even if cars and objects of similar
size can be clearly seen in the 2D satellite images, current dense image matching
methods still have problems in completely reconstructing these objects in terms of
height. However, with increasing size, the 3D reconstruction is very reliable. The
capability and limitation of space-borne very high resolution imagery, specifically
Pléiades (0.70 m) and WorldView-3 (0.31 m) imagery, with respect to the auto-
matic point cloud reconstruction of small isolated objects, was presented in Section
4.6. For this purpose, single buildings, vehicles and trees were analysed. The main
focus was to quantify their detectability in the photogrammetrically derived DSMs,
by estimating their heights as a function of object type and size. The estimated
height was investigated with respect to the following parameters: building length
and width, vehicle length and width and tree crown diameter. As reference, man-
ually measured object heights from the oriented images were used. The results
showed that the DSM-based estimated height of a single object strongly depends on
its size and the effect was quantified accordingly. Starting from very small objects,
which are not elevated against their surroundings and ending with large objects, a
gradual increase of the relative heights was obtained. Objects of a few pixel in size
are hardly mapped in the DSM, whereas, with some dependency on object type,
objects of 12 pixel size are typically mapped with approx. 50% of their real height.
To reach 90% of the real height a minimum object size between 15 and 30 pixel is
necessary. From the performed analyses and investigations, it clearly results that
the vertical accuracy of the DSMs from VHR satellite imagery is influenced by the
acquisition geometry. The Pléiades scenes having a narrow convergence angle of 12°
show a lower vertical performance compared to WorldView-3 scenes, which have a
23° convergence angle on the ground. Nevertheless, even if the acquisition geometry
impacts the DSM accuracy, object height estimation based on automatically derived
DSMs primarily depends on object size, showing a gradual increase of the estimated
heights with object dimensions. Generally, for all single objects (except deciduous
trees) results obtained with WorldView-3 scenes are better than those achieved by
using the Pléiades data. This is because of the very high resolution (0.31 m GSD)
that leads (through photogrammetric processing) to very dense 3D point clouds,
describing the real ground surface and the objects on it more accurately. The dif-
ferent acquisition times also play a significant role in dense image matching quality.
The 3D reconstruction of stable objects, such as buildings, is not affected by the
different acquisition periods, but for deciduous trees, the automatic reconstructed
heights are highly underestimated in the leaf-off season. Their leafless appearance

in the images brings difficulties to the image matching process when finding correct correspondences for tree branches. The results of these investigations - related to the potential of Pléiades and WorldView-3 tri-stereo DSMs for height estimation of individual objects - are published in (Loghin, Otepka-Schremmer, and Pfeifer 2020).

Investigations regarding the visibility analysis from satellite imagery are detailed in Section 4.7. The conducted experiments over Barcelona satellite imagery reveal mathematical geometric connections between visibility in image space and acquisition parameters. The results show that the most important factors affecting the satellite image street visibility are represented by both the city geometry (object heights, street orientations), and satellite viewing geometry (camera azimuth, incidence, viewing, in-track, and cross-track angles). The smaller the incidence, in-track and cross-track angles, the better the street visibility. With the current acquisition geometry characteristics for Barcelona, the highest visibility of 97% is obtained in the forward scene for streets with 45° orientation, while the lowest visibility of only 45% in the backward scene for streets with 135° orientation. To avoid a low visibility and to have as few occlusions as possible, the acquisition collection parameters need to be specified in the technical documentation when ordering satellite images. Another important factor is the acquisition time, for example in the leaf-off periods, to reduce the occlusions caused by deciduous trees. Nevertheless, under these circumstances, the low sun elevation angle leads to a lower image quality .

Aerial or satellite imagery?

For answering this question, the first thing to take into account are the main objectives, conditions and requirements of a project. While both techniques produce digital imagery and offer users a view of the ground from above, aerial imagery is largely used across commercial applications, and satellite imagery for large-scale scientific applications. However, the gap between these two complementary mapping technologies is getting smaller and smaller, as both satellite and aerial imaging systems had great technological improvements over the last years. Due to the available high and very high resolution satellite sensors, the two techniques overlap in terms of spatial resolution, suggesting that other factors like acquisition, accuracy, and processing chain need to be considered for differentiating them.

Since both aerial and satellite imagery possess unique characteristics and strengths, it is important to make a good informed decision about which technology to choose for a specific project requiring geo-spatial data. Looking at the key differences between aerial and satellite imagery, in terms of efficiency, satellites bring a great budget/time cost-benefit, especially for large scale mapping projects where multiple flights for manned or unmanned aircraft are needed. Regarding the project scale, satellite sensors are suited for both small and large scale collections. They collect thousands of square kilometers in minutes, highly overcoming aerial imagery that needs complicated logistics of aircraft flight planning. Other drawbacks of aircraft acquisition are related to land and air restrictions, airspace and travel permits, and planning the suitable take-off and landing points. They are also vulnerable to weather conditions such as heavy winds. None of these issues are encountered by satellite sensors, which can collect isolated, military, conflicted or cross-border locations without any problems. Moreover, ordering satellite imagery is a straightforward fully customisable task, tailored to the user's interests. When tasking, users are allowed to prioritise their areas of interest, to define spectral bands, resolution, collection angles and also the product processing level: basic (1B), standard (OR2A,

2A) and orthorectified (3D, 3A-3X) imagery. Furthermore, the updated real-time weather conditions ensures a cloud-free data acquisition as much as possible, a fact that narrows even more the competitive aspects between aerial data and satellite imagery. The very high resolution of 30 cm make possible to identify objects such as above-ground utility lines (e.g. powerline, pylons) in residential areas, vehicle types, manhole covers, crop species with their boundaries and many more. With a high revisit time, users can trustfully task the exact same area of interest to be collected at specific intervals, which is a crucial feature for automated analysis. The amount of collected data can serve as an important training data source for various programs using machine learning and artificial intelligence to automatically extract and classify features and thereby streamline processing workflows. Furthermore, the already archived imagery can be used to model predictive analytics that are extremely helpful for trend analysis, anomaly detection at a mass scale and profitability estimates (European-Space-Imaging 2020).

Through the present dissertation the level of detail and automation to which the 3D analysis of very high resolution satellite images can advance is shown. The methods introduced and the experiments carried out in this research highlight the potentials and limitations of the automatic 3D information extraction from the VHR optical Pléiades and WorldView-3 stereo and tri-stereo satellite images in areas with different topographic characteristics.

The technology of very high resolution satellite sensors is in a continuous development and the future promises new launches in the next coming years. With their highly-advanced technology and expanded features, the latest sensors will open up new doors for further research investigations and future perspectives.

Chapter 5. Summary and Conclusions

Acknowledgements

Chapter 5.   Summary and Conclusions

Bibliography

Abduelmula, Abdunaser, Bastos, Maria Luisa M, and Gonçalves, José A (2015). "High-accuracy satellite image analysis and rapid DSM extraction for urban environment evaluations (Tripoli-Libya)". In: *International Journal of Environmental and Ecological Engineering* 9.3, pp. 661–666.

Aguilar, Fernando J and Mills, Jon P (2008). "Accuracy assessment of LiDAR-derived digital elevation models". In: *The Photogrammetric Record* 23.122, pp. 148–169.

Aguilar, Manuel A, Aguilar, Fernando J, Agüera, Francisco, and Sánchez, Jaime A (2007). "Geometric accuracy assessment of QuickBird basic imagery using different operational approaches". In: *Photogrammetric Engineering & Remote Sensing* 73.12, pp. 1321–1332.

Aguilar, Manuel A, Aguilar, Fernando J, Mar Saldaña, Maríéa del, Fernández, Ismael, et al. (2012). "Geopositioning accuracy assessment of GeoEye-1 panchromatic and multispectral imagery". In: *Photogrammetric Engineering & Remote Sensing* 78.3, pp. 247–257.

Aguilar, Manuel A, Mar Saldana, Maríéa del, and Aguilar, Fernando J (2013). "Assessing geometric accuracy of the orthorectification process from GeoEye-1 and WorldView-2 panchromatic images". In: *International Journal of Applied Earth Observation and Geoinformation* 21, pp. 427–435.

Aguilar, Manuel A, Nemmaoui, Abderrahim, Aguilar, Fernando J, and Qin, Rongjun (2019). "Quality assessment of digital surface models extracted from WorldView-2 and WorldView-3 stereo pairs over different land covers". In: *GIScience & remote sensing* 56.1, pp. 109–129.

Airbus (2020). *Pléiades Neo - First-rate performance for trusted intelligence. https://www.airbus.com/en/products-services/space/earth-observation/earth-observation-portfolio/pleiades-neo.* Accessed: 2021-11-15.

Alabi, Tunrayo, Haertel, Michael, and Chiejile, Sarah (2016). "Investigating the Use of High Resolution Multi-spectral Satellite Imagery for Crop Mapping in Nigeria-Crop and Landuse Classification using WorldView-3 High Resolution Multispectral Imagery and Landsat8 Data". In: *International Conference on Geographical Information Systems Theory, Applications and Management*. Vol. 2. Scitepress, pp. 109–120.

Alkan, Mehmet, Buyuksalih, Gurcan, Sefercik, Umut G, and Jacobsen, Karsten (2013). "Geometric accuracy and information content of WorldView-1 images". In: *Optical engineering* 52.2, p. 026201.

Allied-Market-Research (2020). *Commercial Satellite Imaging Market. https://www.alliedmarketresearch.com/commercial-satellite-imaging-market.* Accessed: 2021-11-15.

Amberg, V, Dechoz, C, Bernard, L, Greslou, D, De Lussy, F, and Lebegue, L (2013). "In-flight attitude perturbances estimation: Application to PLEIADES-HR satellites". In: *Earth Observing Systems XVIII*. Vol. 8866. SPIE, pp. 327–335.

Arefi, Hossein and Reinartz, Peter (2013). "Building reconstruction using DSM and orthorectified images". In: *Remote Sensing* 5.4, pp. 1681–1703.

Bibliography

Asadzadeh, Saeid and Souza Filho, Carlos Roberto de (2016). "Investigating the capability of WorldView-3 superspectral data for direct hydrocarbon detection". In: *Remote Sensing of Environment* 173, pp. 162–173.

Astrium (2012). *Pléiades Imagery—User Guide v 2.0.* Tech. rep. Toulouse, France: Astrium GEO-Information Services.

Ayoub, François, Leprince, Sébastien, Binet, Renaud, Lewis, Kevin W, Aharonson, Oded, and Avouac, Jean-Philippe (2008). "Influence of camera distortions on satellite image registration and change detection applications". In: *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium.* Vol. 2. IEEE, pp. II–1072.

Bachhofner, Stefan et al. (2020). "Generalized Sparse Convolutional Neural Networks for Semantic Segmentation of Point Clouds Derived from Tri-Stereo Satellite Imagery". In: *Remote Sensing* 12.8, p. 1289.

Bagnardi, Marco, González, Pablo J, and Hooper, Andrew (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.

Balibrea, Mari Paz (2001). "Urbanism, culture and the post-industrial city: Challenging the'Barcelona model'". In: *Journal of Spanish cultural studies* 2.2, pp. 187–210.

Bernard, M, Decluseau, D, Gabet, L, and Nonin, P (2012). "3D capabilities of Pleiades satellite". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39.B3.

Bostater Jr, Charles R, Oney, Taylor S, Rotkiske, Tyler, Aziz, Samin, Morrisette, Charles, Callahan, Kelby, and Mcallister, Devin (2017). "Hyperspectral signatures and WorldView-3 imagery of Indian River Lagoon and Banana River Estuarine water and bottom types". In: *Remote Sensing of the Ocean, Sea Ice, Coastal Waters, and Large Water Regions 2017.* Vol. 10422. International Society for Optics and Photonics, 104220E.

Bostelmann, Jonas and Heipke, Christian (2011). "Modeling spacecraft oscillations in HRSC images of Mars Express". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences; 38-4/W19* 38.W19, pp. 51–56.

Breiman, Leo (2001). "Random forests". In: *Machine learning* 45.1, pp. 5–32.

Breiman, Leo, Friedman, Jerome, Stone, Charles J, and Olshen, Richard A (1984). *Classification and regression trees.* CRC press.

Bühler, Yves, Marty, Mauro, and Ginzler, Christian (2012). "High resolution DEM generation in high-alpine terrain using airborne remote sensing techniques". In: *Transactions in GIS* 16.5, pp. 635–647.

Burrard, Sidney Gerald (1907). *A Sketch of the Geography and Geology of the Himalaya Mountains and Tibet.* Vol. 1. Superintendent government printing, India.

Caglar, Bayik, Becek, Kazimierz, Mekik, C, and Ozendi, M (2018). "On the vertical accuracy of the ALOS world 3D-30m digital elevation model". In: *Remote Sensing Letters* 9.6, pp. 607–615.

Cao, Jinshan, Fu, Jianhong, Yuan, Xiuxiao, and Gong, Jianya (2017). "Nonlinear bias compensation of ZiYuan-3 satellite imagery with cubic splines". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 133, pp. 174–185.

Cao, Yinxia and Huang, Xin (2021). "A deep learning method for building height estimation using high-resolution multi-view imagery over urban areas: A case study of 42 Chinese cities". In: *Remote Sensing of Environment* 264, p. 112590.

Comp, C and Mulawa, D (2015). *WorldView-3 Geometric Calibration*. Tech. rep. Florida, USA: Digital Globe.

Contreras, Marco A, Staats, Wesley, Yiang, Jian, and Parrott, David (2017). "Quantifying the Accuracy of LiDAR-Derived DEM in Deciduous Eastern Forests of the Cumberland Plateau". In: *Journal of Geographic Information System* 9.3, pp. 339–353.

Cook, Joel (1910). *The Mediterranean and Its Borderlands*.

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

d'Angelo, Pablo, Lehner, Manfred, Krauss, Thomas, Hoja, Danielle, and Reinartz, Peter (2008). "Towards automated DEM generation from high resolution stereo satellite images". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, International Society for Photogrammetry and Remote Sensing* 37.B4, pp. 1137–1342.

d'Angelo, Pablo and Reinartz, Peter (2011). "Semiglobal matching results on the ISPRS stereo matching benchmark". In:

De Deuge, Mark, Quadros, Alastair, Hung, Calvin, and Douillard, Bertrand (2013). "Unsupervised feature learning for classification of outdoor 3d scans". In: *Australasian Conference on Robitics and Automation*. Vol. 2, p. 1.

De Geer, Sten (1923). "Greater Stockholm: a geographical interpretation". In: *Geographical Review* 13.4, pp. 497–506.

De Lussy, Françoise, Greslou, Daniel, Dechoz, Cécile, Amberg, Virginie, Delvit, Jean Marc, Lebegue, Laurent, Blanchet, Gwendoline, and Fourest, Sebastien (2012). "Pleiades HR in flight geometrical calibration: Location and mapping of the focal plane". In: *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci* 39, pp. 519–523.

Delgado, Maria M (2012). "Performing Barcelona: Cultural Tourism, Geography and Identity". In: University of Wales Press.

Dial, G and Grodecki, J (2004). "Test ranges for metric calibration and validation of high-resolution satellite imaging systems". In: *Post-Launch Calibration of Satellite Sensors: Proceedings of the International Workshop on Radiometric and Geometric Calibration, December 2003, Mississippi, USA*. Vol. 2. CRC Press, p. 171.

Dial, Gene and Grodecki, Jacek (2002). "Block adjustment with rational polynomial camera models". In: *Proceedings of ASPRS 2002 Conference, Washington, DC*, pp. 22–26.

DigitalGlobe (2012). *DigitalGlobe Core Imagery Products Guide*. Tech. rep. Toulouse, France: DigitalGlobe.

Dong, Yang, Lei, Rong, Fan, Dazhao, Gu, Linyu, and Ji, Song (2020). "A NOVEL RPC BIAS MODEL FOR IMPROVING THE POSITIONING ACCURACY OF SATELLITE IMAGES." In: *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 5.2.

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

Bibliography

eoPortal (2021). *eoPortal Directory - WorldView Legion Constellation*. https://
directory.eoportal.org/web/eoportal/satellite-missions/v-w-x-y-z/worldview-
legion. Accessed: 2021-11-15.

ESA (2021). *Earth online - Pléiades Neo mission*. https://earth.esa.int/eogateway/
missions/pleiades-neo. Accessed: 2021-11-15.

Esch, Thomas et al. (2020). "Towards a large-scale 3D modeling of the built envi-
ronment—joint analysis of TanDEM-X, Sentinel-2 and open street map data". In:
Remote Sensing 12.15, p. 2391.

European-Space-Imaging (2020). *30 cm Satellite Imagery as an Alternative to Aerial
Data*. https://www.euspaceimaging.com/30-cm-satellite-imagery-as-an-
alternative-to-aerial-data/. Accessed: 2021-12-15.

Flamanc, D and Maillet, G (2005). "Evaluation of 3D city model production from
PLEIADES HR satellite images and 2D ground maps". In: *Urban, Tempe, USA*.

Förstner, Wolfgang (1982). "On the geometric precision of digital correlation".
In: *Proceedings of the Symposium of the ISPRS Commision III Helsinki, 1982*,
pp. 176–189.

– (1984). "Quality assessment of object location and point transfer using digital
image correlation techniques". In: *IBID* 25.3, pp. 197–219.

– (1986). "A feature based correspondence algorithm for image matching". In: *IS-
PRS ComIII, Rovaniemi*, pp. 150–166.

Fraser, Clive S, Dial, G, and Grodecki, J (2006). "Sensor orientation via RPCs". In:
ISPRS journal of Photogrammetry and Remote Sensing 60.3, pp. 182–194.

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

– (2005). "Bias-compensated RPCs for sensor orientation of high-resolution satellite
imagery". In: *Photogrammetric Engineering & Remote Sensing* 71.8, pp. 909–915.

Geiß, Christian, Schrade, Henrik, Pelizari, Patrick Aravena, and Taubenböck, Hannes
(2020). "Multistrategy ensemble regression for mapping of built-up density and
height with Sentinel-2 data". In: *ISPRS Journal of Photogrammetry and Remote
Sensing* 170, pp. 57–71.

Gerke, Markus and Xiao, Jing (2013). "Supervised and unsupervised MRF based 3D
scene classification in multiple view airborne oblique images". In: *ISPRS Annals
of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2.3,
pp. 25–30.

Gharib Bafghi, Zeinab, Tian, Jiaojiao, d'Angelo, Pablo, and Reinartz, Peter (2016).
"A new algorithm for void filling in a DSM from stereo satellite images in urban
areas". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial
Information Sciences, 2016* 3, pp. 55–61.

Gibbins, Danny and Swierkowski, Leszek (2009). "A comparison of terrain classifica-
tion using local feature measurements of 3-dimensional colour point-cloud data".
In: *2009 24th International conference image and vision computing New Zealand*.
IEEE, pp. 293–298.

Greslou, D, De Lussy, F, Delvit, JM, Dechoz, C, and Amberg, V (2012). "Pleiades-
HR innovative techniques for geometric image quality commissioning". In: *Int.
Arch. Photogramm. Remote Sens. Spat. Inf. Sci*, pp. 543–547.

Griffiths, Tony (2009). *Stockholm: A cultural history*. Oxford University Press.

Grilli, Eleonora, Menna, Fabio, and Remondino, Fabio (2017). "A review of point clouds segmentation and classification algorithms". In: *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 42, p. 339.

Grodecki, Jacek and Dial, Gene (2002). "IKONOS geometric accuracy validation". In: *International archives of photogrammetry remote sensing and spatial information sciences* 34.1, pp. 50–55.

– (2003). "Block adjustment of high-resolution satellite images described by rational polynomials". In: *Photogrammetric Engineering & Remote Sensing* 69.1, pp. 59–68.

Grohman, Greg, Kroenung, George, and Strebeck, John (2006). "Filling SRTM voids: The delta surface fill method". In: *Photogrammetric Engineering and Remote Sensing* 72.3, pp. 213–216.

Gunter (2020). *WorldView-Scout 1, 2, 3, 4, 5, 6. https://space.skyrocket.de/doc_sdat/worldview-scout.htm*. Accessed: 2021-11-15.

Guo, Yulan, Wang, Hanyun, Hu, Qingyong, Liu, Hao, Liu, Li, and Bennamoun, Mohammed (2020). "Deep learning for 3d point clouds: A survey". In: *IEEE transactions on pattern analysis and machine intelligence.*

Gwinner, Klaus et al. (2010). "Topography of Mars from global mapping by HRSC high-resolution digital terrain models and orthoimages: Characteristics and performance". In: *Earth and Planetary Science Letters* 294.3-4, pp. 506–519.

Haala, Norbert and Rothermel, Mathias (2012). "Dense multi-stereo matching for high quality digital elevation models". In: *PFG Photogrammetrie, Fernerkundung, Geoinformation*, pp. 331–343.

Habib, Ayman, Ghanma, Mwafag, Morgan, Michel, and Al-Ruzouq, Rami (2005). "Photogrammetric and LiDAR data registration using linear features". In: *Photogrammetric Engineering & Remote Sensing* 71.6, pp. 699–707.

Hackel, Timo, Savinov, Nikolay, Ladicky, Lubor, Wegner, Jan D, Schindler, Konrad, and Pollefeys, Marc (2017). "Semantic3d. net: A new large-scale point cloud classification benchmark". In: *arXiv preprint arXiv:1704.03847.*

Heipke, Christian (1992). "A global approach for least-squares image matching and surface reconstruction in object space". In: *Photogrammetric engineering and remote sensing* 58, pp. 317–317.

– (1997). "Automation of interior, relative, and absolute orientation". In: *ISPRS journal of photogrammetry and remote sensing* 52.1, pp. 1–19.

Hirschmuller, Heiko (2007). "Stereo processing by semiglobal matching and mutual information". In: *IEEE Transactions on pattern analysis and machine intelligence* 30.2, pp. 328–341.

Hobi, Martina L and Ginzler, Christian (2012). "Accuracy assessment of digital surface models based on WorldView-2 and ADS80 stereo remote sensing data". In: *Sensors* 12.5, pp. 6347–6368.

Hoffmann, Andrea and Lehmann, Frank (2000). "Vom Mars zur Erde—die erste digitale Orthobildkarte Berlin mit Daten der Kamera HRSC-A". In: *KN-Journal of Cartography and Geographic Information* 50.2, pp. 61–72.

Höfle, Bernhard, Hollaus, Markus, and Hagenauer, Julian (2012). "Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 67, pp. 134–147.

Bibliography

Hoja, Danielle, Schneider, Mathias, Müller, Rupert, Lehner, Manfred, and Reinartz, Peter (2008). "Comparison of orthorectification methods suitable for rapid mapping using direct georeferencing and RPC for optical satellite data". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37.B4, pp. 1617–1624.

Hong, Zhonghua, Tong, Xiaohua, Liu, Shijie, Chen, Peng, Xie, Huan, and Jin, Yanmin (2015). "A comparison of the performance of bias-corrected RSMs and RFMs for the geo-positioning of high-resolution satellite stereo imagery". In: *Remote Sensing* 7.12, pp. 16815–16830.

Hu, F, Gao, XM, Li, GY, and Li, M (2016). "DEM Extraction from WorldView-3 Stereo-Images and Accuracy Evaluation". In: *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 41.

Hu, WenYuan, Yang, GengYin, and Yuan, Hui (2008). "Application and accuracy evaluation of Leica ADS40 for large scale mapping". In: *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci* 37, pp. 605–610.

Hu, Yong, Tao, Vincent, and Croitoru, Arie (2004). "Understanding the rational function model: methods and applications". In: *International archives of photogrammetry and remote sensing* 20.6, pp. 119–124.

Inglada, Jordi and Christophe, Emmanuel (2009). "The Orfeo Toolbox remote sensing image processing software". In: *2009 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 4. IEEE, pp. IV–733.

Iwasaki, Akira (2011). "Detection and estimation satellite attitude jitter using remote sensing imagery". In: *Advances in Spacecraft Technologies* 13, pp. 257–272.

Jacobsen, K (2018). "Systematic geometric image errors of very high resolution optical satellites". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives 42 (2018), Nr. 1* 42.1, pp. 233–238.

Jacobsen, K et al. (2008). "Satellite image orientation". In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37, pp. 703–709.

Jacobsen, Karsten (2005). "Geometry of satellite images-calibration and mathematical models". In: *Proceedings of the KSRS Conference*. The Korean Society of Remote Sensing, pp. 182–185.

– (n.d.). "Verbesserung der Geometrie von Satellitenbildern durch Höhenmodelle". In: *Publikationen der Deutschen Gesellschaft für Photogrammetrie* (), p. 13.

Jacobsen, Karsten and Topan, H (2015). "DEM generation with short base length Pleiades triplet". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives 40 (2015), Nr. 3W2* 40.3W2, pp. 81–86.

Janson, Carl-Gunnar and Rudolfsson, Bengt (1965). "Ecological and geographical distances in Stockholm". In: *Acta Sociologica* 8.4, pp. 285–292.

Jiang, Yong-hua, Zhang, Guo, Tang, Xinming, Li, Deren, and Huang, Wen-chao (2014). "Detection and correction of relative attitude errors for ZY1-02C". In: *IEEE Transactions on Geoscience and Remote Sensing* 52.12, pp. 7674–7683.

Jiang, Yong-hua, Zhang, Guo, Tang, Xinming, Li, De-ren, Wang, Taoyang, Huang, Wen-chao, and Li, Li-tao (2015). "Improvement and assessment of the geometric accuracy of Chinese high-resolution optical satellites". In: *IEEE Journal of Se-*

lected Topics in Applied Earth Observations and Remote Sensing 8.10, pp. 4841–4852.

Kim, Taejung, Javzandulam, Ts, and Lee, Tae-Yoon (2007). "Semiautomatic reconstruction of building height and footprints from single satellite images". In: *2007 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 4737–4740.

Koenig, Jan and Gueguen, Lionel (2016). "A comparison of land use land cover classification using superspectral WorldView-3 vs hyperspectral imagery". In: *2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (Whispers)*. IEEE, pp. 1–5.

Kornus, Wolfgang, Lehner, Manfred, and Schroeder, Manfred (1999). "Geometric inflight-calibration by block adjustment using MOMS-2P-imagery of three intersecting stereo-strips". In: *Proceedings of the ISPRS Workshop on Sensors and Mapping from Space*, pp. 27–30.

Kraus, Karl (2012). *Photogrammetrie: Geometrische Informationen aus Photographien und Laserscanneraufnahmen*. Walter de Gruyter.

Kuester, Michele (2016). "Radiometric use of worldview-3 imagery". In: *DigitalGlobe: Longmont, CO, USA*.

Lafarge, Florent, Descombes, Xavier, Zerubia, Josiane, and Pierrot-Deseilligny, Marc (2006). "An automatic building reconstruction method: a structural approach using high resolution satellite images". In: *2006 International Conference on Image Processing*. IEEE, pp. 1205–1208.

Lehner, Manfred and Müller, Rupert (2003). "Quality check of MOMS-2P ortho-images of semi-arid landscapes". In: *Proc. ISPRS Workshop High Resolution Mapping Space*, pp. 1–5.

Leotta, Matthew J et al. (2019). "Urban semantic 3d reconstruction from multiview satellite imagery". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, p. 10.

Li, Nan, Kähler, Olaf, and Pfeifer, Norbert (2021). "A comparison of deep learning methods for airborne lidar point clouds classification". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, pp. 6467–6486.

Li, Rongxing, Hwangbo, Juwon, Chen, Yunhang, and Di, Kaichang (2011). "Rigorous photogrammetric processing of HiRISE stereo imagery for Mars topographic mapping". In: *IEEE Transactions on Geoscience and Remote Sensing* 49.7, pp. 2558–2572.

Liu, Xiaoye, Zhang, Zhenyu, Peterson, Jim, and Chandra, Shobhit (2007). "LiDAR-derived high quality ground control information and DEM for image orthorectification". In: *GeoInformatica* 11.1, pp. 37–53.

Loghin, Ana-Maria, Oniga, Valeria-Ersilia, and Giurma-Handley, C (2018). "3D Point Cloud Classification of Natural Environments Using Airborne Laser Scanning Data". In: *American Journal of Engineering Research (AJER)* 7.2, pp. 191–197.

Loghin, Ana-Maria, Otepka-Schremmer, Johannes, Karel, Wilfried, Pöchtrager, Markus, and Pfeifer, Norbert (2019). "Accuracy Analysis of Digital Elevation Models from very High Resolution Satellite Imagery". In: *Proceedings of Dreiländertagung der DGPF, der OVG und der SGPF - Publikationen der DGPF*. Vol. 28, pp. 123–137.

Bibliography

Loghin, Ana-Maria, Otepka-Schremmer, Johannes, and Pfeifer, Norbert (2020). "Potential of Pléiades and WorldView-3 Tri-Stereo DSMs to Represent Heights of Small Isolated Objects". In: *Sensors* 20.9, p. 2695.

Loghin, Ana-Maria, Otepka-Schremmer, Johannes, Ressl, Camillo, and Pfeifer, Norbert (2022). "Improvement of VHR Satellite Image Geometry with High Resolution Elevation Models". In: *Remote Sensing* 14.10, p. 2303. ISSN: 2072-4292. DOI: 10.3390/rs14102303. URL: https://www.mdpi.com/2072-4292/14/10/2303.

Loghin, Ana-Maria, Pfeifer, Norbert, and Otepka-Schremmer, Johannes (2020). "Supervised Classification and its Repeatability for Point Clouds from Dense Vhr Tri-Stereo Satellite Image Matching Using Machine Learning". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2, pp. 525–532.

Long, Jonathan, Shelhamer, Evan, and Darrell, Trevor (2015). "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440.

Maas, Hans-Gerd (2000). "Least-squares matching with airborne laserscanning data in a TIN structure". In: *International Archives of Photogrammetry and Remote Sensing* 33.B3/1; PART 3, pp. 548–555.

Mahendra, Anjali and Seto, Karen C (2019). "Upward and outward growth: managing urban expansion for more equitable cities in the global south". In:

Mallet, Clément, Bretar, Frédéric, Roux, Michel, Soergel, Uwe, and Heipke, Christian (2011). "Relevance assessment of full-waveform lidar data for urban area classification". In: *ISPRS journal of photogrammetry and remote sensing* 66.6, S71–S84.

Mandanici, Emanuele, Girelli, Valentina A, and Poluzzi, Luca (2019). "Metric accuracy of digital elevation models from worldview-3 stereo-pairs in urban areas". In: *Remote Sensing* 11.7, p. 878.

Marshall, Tim (2004). *Transforming Barcelona*. Psychology Press.

Mattson, S, Boyd, A, Kirk, RL, Cook, DA, and Howington-Kraus, E (2009). "Hi-JACK: Correcting spacecraft jitter in HiRISE images of Mars". In: *Health Manag. Technol* 33.5, A162.

Mattson, S, Robinson, M, et al. (2010). "Early assessment of spacecraft jitter in LROC-NAC". In: *41st Annual Lunar and Planetary Science Conference*. Vol. 41, p. 1871.

Maxar (2020). *WorldView Legion. https://www.maxar.com/splash/it-takes-a-legion*. Accessed: 2021-11-15.

Modiri, M, Masumi, M, and Eftekhari, A (2015). "Automatic classification of point clouds extracted from Ultracam stereo images." In: *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 40.

Moreira, JM Macay, Nex, F, Agugiaro, G, Remondino, F, and Lim, Nancy J (2013). "From DSM to 3D building models: a quantitative evaluation". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1, W1.

Mumtaz, Rafia and Palmer, Phil (2013). "Attitude determination by exploiting geometric distortions in stereo images of DMC camera". In: *IEEE Transactions on Aerospace and Electronic Systems* 49.3, pp. 1601–1625.

Munoz, Daniel, Bagnell, J Andrew, Vandapel, Nicolas, and Hebert, Martial (2009). "Contextual classification with functional max-margin markov networks". In: *2009*

IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 975–982.

Niemeyer, Joachim, Rottensteiner, Franz, and Soergel, Uwe (2014). "Contextual classification of lidar data and building object detection in urban areas". In: *ISPRS journal of photogrammetry and remote sensing* 87, pp. 152–165.

Nikolakopoulos, Konstantinos G and Lathourakis, George (2005). "Along the track vs across the track satellite stereo-pair for DTM creation". In: *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05.* Vol. 8. IEEE, pp. 5324–5327.

Noguchi, Mayumi, Fraser, Clive S, Nakamura, Takayuki, Shimono, Takahiro, and Oki, Shoichi (2004). "Accuracy assessment of QuickBird stereo imagery". In: *The Photogrammetric Record* 19.106, pp. 128–137.

Oh, Jaehong and Lee, Changno (2015). "Automated bias-compensation of rational polynomial coefficients of high resolution satellite imagery based on topographic maps". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 100, pp. 14–22.

Otepka-Schremmer, Johannes, Pöchtrager, Markus, Holaus, Markus, Loghin, Ana-Maria, Piermattei, Livia, Briese, Christian, and Pfeifer, Norbert (2019). "Landesweite Punktwolkenprozessierung mit OPALS am Supercomputer VSC3". In: *Proceedings of 20. Internationale Geodätische Woche Obergurgl*, pp. 165–176.

Otepka, Johannes, Ghuffar, Sajid, Waldhauser, Christoph, Hochreiter, Ronald, and Pfeifer, Norbert (2013). "Georeferenced point clouds: A survey of features and point cloud management". In: *ISPRS International Journal of Geo-Information* 2.4, pp. 1038–1065.

Pan, Hongbo, Zhang, Guo, Tang, Xinming, Li, Deren, Zhu, Xiaoyong, Zhou, Ping, and Jiang, Yonghua (2013). "Basic products of the ZiYuan-3 satellite and accuracy evaluation". In: *Photogrammetric Engineering & Remote Sensing* 79.12, pp. 1131–1145.

Pan, Jun, Che, Chengbang, Zhu, Ying, and Wang, Mi (2017). "Satellite jitter estimation and validation using parallax images". In: *Sensors* 17.1, p. 83.

Panagiotakis, Emmanouil, Chrysoulakis, Nektarios, Charalampopoulou, Vasiliki, and Poursanidis, Dimitris (2018). "Validation of Pleiades Tri-Stereo DSM in urban areas". In: *ISPRS International Journal of Geo-Information* 7.3, p. 118.

Panem, C, Bignalet-Cazalet, F, and Baillarin, S (2012). "Pleiades-HR system products performance after in-orbit commissioning phase". In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39, B1.

Partovi, Tahmineh, Fraundorfer, Friedrich, Bahmanyar, Reza, Huang, Hai, and Reinartz, Peter (2019). "Automatic 3-D building model reconstruction from very high resolution stereo satellite imagery". In: *Remote Sensing* 11.14, p. 1660.

Perko, Roland, Raggam, Hannes, Gutjahr, Karlheinz, and Schardt, Mathias (2014). "Assessment of the mapping potential of Pléiades stereo and triplet data". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2.3, p. 103.

Perko, Roland, Raggam, Hannes, Schardt, Mathias, and Roth, Peter Michael (2018). "Very high resolution mapping with the Pleiades satellite constellation". In: *Am. J. Remote Sens* 6.89, p. 2019.

Bibliography

Pfeifer, Norbert, Mandlburger, Gottfried, Otepka, Johannes, and Karel, Wilfried (2014). "OPALS–A framework for Airborne Laser Scanning data analysis". In: *Computers, Environment and Urban Systems* 45, pp. 125–136.

Piermattei, Livia, Marty, Mauro, Ginzler, Christian, Pöchtrager, Markus, Karel, Wilfried, Ressl, Camillo, Pfeifer, Norbert, and Hollaus, Markus (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.

Piermattei, Livia, Marty, Mauro, Karel, Wilfried, Ressl, Camillo, Hollaus, Markus, Ginzler, Christian, and Pfeifer, Norbert (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.

Pierrot-Deseilligny, Marc and Paparoditis, Nicolas (2006). "A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery". In: *Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36.1/W41, pp. 1–5.

Piltz, Björn, Bayer, Steven, and Poznanska, Anna-Maria (2016). "Volume based DTM generation from very high resolution photogrammetric DSMs". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 41, pp. 83–90.

Poli, D, Remondino, F, Angiuli, E, and Agugiaro, G (2013). "Evaluation of Pléiades-1A triplet on Trento testfield". In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-1 W* 1, pp. 287–292.

Poli, D, Remondino, Fabio, Angiuli, E, and Agugiaro, G (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.

Poli, Daniela and CARAVAGGI, Ivano (2012). "Digital surface modelling and 3D information extraction from spaceborne very high resolution stereo pairs". In: *JRC Scientific and Technical Reports, Ispra*.

Poli, Daniela and Caravaggi, Ivano (2013). "3D modeling of large urban areas with stereo VHR satellite imagery: lessons learned". In: *Natural hazards* 68.1, pp. 53–78.

Poli, Daniela and Toutin, Thierry (2012). "Review of developments in geometric modelling for high resolution satellite pushbroom sensors". In: *The Photogrammetric Record* 27.137, pp. 58–73.

Prakash, PS and Aithal, Bharath H (2021). "Digital Building-Height Preparation from Satellite Stereo Images". In: *Photogrammetric Engineering & Remote Sensing* 87.8, pp. 557–566.

Pratt, John Henry (1855). "I. On the attraction of the Himalaya Mountains, and of the elevated regions beyond them, upon the plumb-line in India". In: *Philosophical Transactions of the Royal Society of London* 145, pp. 53–100.

Rau, Jiann-Yeou, Jhan, Jyun-Ping, and Hsu, Ya-Ching (2014). "Analysis of oblique aerial images for land cover and point cloud classification in an urban environment". In: *IEEE Transactions on Geoscience and Remote Sensing* 53.3, pp. 1304–1319.

Remondino, F, Spera, Maria Grazia, Nocerino, Erica, Menna, Fabio, and Nex, Francesco (2014). "State of the art in high density image matching". In: *The photogrammetric record* 29.146, pp. 144–166.

Ressl, Camillo, Brockmann, Herbert, Mandlburger, Gottfried, and Pfeifer, Norbert (2016). "Dense Image Matching vs. Airborne Laser Scanning–Comparison of two methods for deriving terrain models". In: *Photogrammetrie-Fernerkundung-Geoinformation* 2016.2, pp. 57–73.

Ressl, Camillo, Kager, Helmut, and Mandlburger, Gottfried (2008). "Quality checking of ALS projects using statistics of strip differences". In: *International Archives of Photogrammetry and Remote Sensing* 37.B3b, pp. 253–260.

Robertson, Brian C (2003). "Rigorous geometric modeling and correction of QuickBird imagery". In: *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477)*. Vol. 2. IEEE, pp. 797–802.

Rottensteiner, Franz, Sohn, Gunho, Jung, Jaewook, Gerke, Markus, Baillard, Caroline, Benitez, Sebastien, and Breitkopf, Uwe (2012). "The ISPRS benchmark on urban object classification and 3D building reconstruction". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3 (2012), Nr. 1* 1.1, pp. 293–298.

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

Sadeghian, Saeid, Zoej, Mohammad Javad Valadan, Delavar, Mahmoud Reza, and Abootalebi, Ahmad (2001). "Precision rectification of high resolution satellite imagery without ephemeris data". In: *International Journal of Applied Earth Observation and Geoinformation* 3.4, pp. 366–371.

Santillan, Jojene R, Makinano-Santillan, Meriam, and Makinano, Ronald M (2016). "Vertical accuracy assessment of ALOS World 3D-30M Digital Elevation Model over northeastern Mindanao, Philippines". In: *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, pp. 5374–5377.

Schenk, Toni, Krupnik, Amnon, and Postolov, Yuri (2000). "Comparative study of surface matching algorithms". In: *International Archives of Photogrammetry and Remote Sensing* 33.B4, pp. 518–524.

Schwind, Peter, Schneider, Mathias, Palubinskas, Gintautas, Storch, Tobias, Muller, Rupert, and Richter, Rudolf (2009). "Processors for ALOS optical data: Deconvolution, DEM generation, orthorectification, and atmospheric correction". In: *IEEE transactions on geoscience and remote sensing* 47.12, pp. 4074–4082.

Sefercik, Umut G, Alkan, Mehmet, Atalay, Can, Jacobsen, Karsten, Büyüksalih, Gürcan, and Karakış, Serkan (2020). "Optimizing the Achievable Information Content Extraction from WorldView-4 Stereo Imagery". In: *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 88.6, pp. 449–461.

Shen, Xiang, Li, Qingquan, Wu, Guofeng, and Zhu, Jiasong (2017). "Bias compensation for rational polynomial coefficients of high-resolution satellite imagery by local polynomial modeling". In: *Remote Sensing* 9.3, p. 200.

Shen, Xiang, Liu, Bin, and Li, Qing-Quan (2017). "Correcting bias in the rational polynomial coefficients of satellite imagery using thin-plate smoothing splines". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 125, pp. 125–131.

Sirmacek, Beril, Taubenbock, Hannes, Reinartz, Peter, and Ehlers, Manfred (2012). "Performance evaluation for 3-D city model generation of six different DSMs from air-and spaceborne sensors". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5.1, pp. 59–70.

Bibliography

Stentoumis, Christos, Karkalou, E, and Karras, G (2015). "A review and evaluation of penalty functions for Semi-Global Matching". In: *2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*. IEEE, pp. 167–172.

Stumpf, A, Malet, J-P, Allemand, P, and Ulrich, P (2014). "Surface reconstruction and landslide displacement measurements with Pléiades satellite images". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 95, pp. 1–12.

Szeliski, Richard (2010). *Computer vision: algorithms and applications*. Springer Science & Business Media.

Tadono, Takeo, Takaku, Junichi, Tsutsui, Ken, Oda, Fumiko, and Nagai, Hiroto (2015). "Status of "ALOS World 3D (AW3D)" global DSM generation". In: *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, pp. 3822–3825.

Takaku, Junichi and Tadono, Takeo (2010). "High resolution dsm generation from alos prism-processing status and influence of attitude fluctuation". In: *2010 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 4228–4231.

Takaku, Junichi, Tadono, Takeo, and Tsutsui, Ken (2014). "GENERATION OF HIGH RESOLUTION GLOBAL DSM FROM ALOS PRISM." In: *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 2.4.

Tan, Pang-Ning, Steinbach, Michael, and Kumar, Vipin (2005). *Introduction to data mining (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA. ISBN0321321367.

Tao, C Vincent and Hu, Yong (2001). "A comprehensive study of the rational function model for photogrammetric processing". In: *Photogrammetric engineering and remote sensing* 67.12, pp. 1347–1358.

Tao, C Vincent, Hu, Yong, and Jiang, W (2004). "Photogrammetric exploitation of IKONOS imagery for mapping applications". In: *International Journal of Remote Sensing* 25.14, pp. 2833–2853.

Teo, Tee-Ann (2011). "Bias compensation in a rigorous sensor model and rational function model for high-resolution satellite images". In: *Photogrammetric Engineering & Remote Sensing* 77.12, pp. 1211–1220.

Teshima, Yu and Iwasaki, Akira (2007). "Correction of attitude fluctuation of Terra spacecraft using ASTER/SWIR imagery with parallax observation". In: *IEEE Transactions on Geoscience and Remote Sensing* 46.1, pp. 222–227.

Therneau, Terry M, Atkinson, Elizabeth J, et al. (1997). *An introduction to recursive partitioning using the RPART routines*. Tech. rep. Technical report Mayo Foundation.

Tong, Xiaohua, Li, Lingyun, Liu, Shijie, Xu, Yusheng, Ye, Zhen, Jin, Yanmin, Wang, Fengxiang, and Xie, Huan (2015). "Detection and estimation of ZY-3 three-line array image distortions caused by attitude oscillation". In: *ISPRS journal of photogrammetry and remote sensing* 101, pp. 291–309.

Tong, Xiaohua, Liu, Shijie, and Weng, Qihao (2010). "Bias-corrected rational polynomial coefficients for high accuracy geo-positioning of QuickBird stereo imagery". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.2, pp. 218–226.

Tong, Xiaohua, Ye, Zhen, et al. (2014). "Framework of jitter detection and compensation for high resolution satellites". In: *Remote Sensing* 6.5, pp. 3944–3964.

Topan, Hüseyin, Oruc, Murat, Taşkanat, Talha, and Cam, Ali (2013). "Combined efficiency of RPC and DEM accuracy on georeferencing accuracy of orthoimage:

Case study with Pléiades panchromatic mono image". In: *IEEE Geoscience and Remote Sensing Letters* 11.6, pp. 1148–1152.

Toutin, Thierry (2004). "Geometric processing of remote sensing images: models, algorithms and methods". In: *International journal of remote sensing* 25.10, pp. 1893–1924.

– (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.

Toutin, Thierry and Cheng, Philip (2002). "Comparison of automated digital elevation model extraction results using along-track ASTER and across-track SPOT stereo images". In: *Optical Engineering-Bellingham- International Society for Optical Engineering* 41.9, pp. 2102–2106.

Tran, Thi Huong Giang, Otepka, Johannes, Wang, Di, and Pfeifer, Norbert (2018). "Classification of image matching point clouds over an urban area". In: *International Journal of Remote Sensing* 39.12, pp. 4145–4169.

Trimble (2016a). *Match-AT Reference Manual*. Tech. rep. Sunnyvale, CA, USA: Trimble Inc.

– (2016b). *Match-T DSM Reference Manual*. Tech. rep. Sunnyvale, CA, USA: Trimble Inc.

Vallet, Bruno, Brédif, Mathieu, Serna, Andrés, Marcotegui, Beatriz, and Paparoditis, Nicolas (2015). "TerraMobilita/iQmulus urban point cloud analysis benchmark". In: *Computers & Graphics* 49, pp. 126–133.

Vosselman, George and Maas, Hans-Gerd (2010). *Airborne and terrestrial laser scanning*. CRC press.

Waldhauser, Christoph, Hochreiter, Ronald, Otepka, Johannes, Pfeifer, Norbert, Ghuffar, Sajid, Korzeniowska, Karolina, and Wagner, Gerald (2014). "Automated classification of airborne laser scanning point clouds". In: *Solving Computationally Expensive Engineering Problems*. Springer, pp. 269–292.

Wang, Mi, Zhu, Ying, Pan, Jun, Yang, Bo, and Zhu, Quansheng (2016). "Satellite jitter detection and compensation using multispectral imagery". In: *Remote Sensing Letters* 7.6, pp. 513–522.

Wenzel, Konrad, Rothermel, Mathias, Haala, Norbert, and Fritsch, Dieter (2013). "SURE–The ifp software for dense image matching". In: *Photogrammetric week*. Vol. 13, pp. 59–70.

Xie, Yuxing, Tian, Jiaojiao, and Zhu, Xiao Xiang (2019). "A review of point cloud semantic segmentation". In: *arXiv preprint arXiv:1908.08854*.

Xu, Bo, Zhang, Xu, Li, Zhixin, Leotta, Matt, Chang, Shih-Fu, and Shan, Jie (2020). "Deep Learning Guided Building Reconstruction from Satellite Imagery-derived Point Clouds". In: *arXiv preprint arXiv:2005.09223*.

Yang, Qinye and Zheng, Du (2004). *Tibetan geography*. Wuzhou Communication Publishing House.

Yi, Yaning, Zhang, Zhijie, Zhang, Wanchang, Zhang, Chuanrong, Li, Weidong, and Zhao, Tian (2019). "Semantic segmentation of urban buildings from vhr remote sensing imagery using a deep convolutional neural network". In: *Remote Sensing* 11.15, p. 1774.

Zambanini, Sebastian, Loghin, Ana-Maria, Pfeifer, Norbert, Soley, Elena Màrmol, and Sablatnig, Robert (2020). "Detection of parking cars in stereo satellite images". In: *Remote Sensing* 12.13, p. 2170.

Bibliography

Zhang, Li (2005). *Automatic digital surface model (DSM) generation from linear array images*. ETH Zurich.

Zhang, Yongjun, Lu, Yihui, Wang, Lei, and Huang, Xu (2011). "A new approach on optimization of the rational function model of high-resolution satellite imagery". In: *IEEE Transactions on Geoscience and Remote Sensing* 50.7, pp. 2758–2764.

Zhang, Yongjun, Zheng, Maoteng, Xiong, Xiaodong, and Xiong, Jinxin (2014). "Multistrip bundle block adjustment of ZY-3 satellite imagery by rigorous sensor model without ground control point". In: *IEEE Geoscience and Remote Sensing Letters* 12.4, pp. 865–869.

Zhang, Zhenxin, Zhang, Liqiang, Tong, Xiaohua, Mathiopoulos, P Takis, Guo, Bo, Huang, Xianfeng, Wang, Zhen, and Wang, Yuebin (2016). "A multilevel point-cluster-based discriminative feature for ALS point cloud classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.6, pp. 3309–3321.

Zhou, Yu, Parsons, Barry, Elliott, John R, Barisin, Ivana, and Walker, Richard T (2015). "Assessing the ability of Pleiades stereo imagery to determine height changes in earthquakes: A case study for the El Mayor-Cucapah epicentral area". In: *Journal of Geophysical Research: Solid Earth* 120.12, pp. 8793–8808.