Creating and evaluating Digital Elevation Models from Satellite Imagery

Master thesis Geoscience and Remote Sensing

Robin Claesen

Delft University of Technology

TU Delft

CGI

Creating and evaluating Digital Elevation Models from Satellite Imagery

Master thesis Geoscience and Remote Sensing

by

Robin Claesen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday July 16, 2024 at 14:30 AM.
Student Number: 4709977

Thesis committee	Dr. R.C. Lindenbergh	TU Delft, supervisor
	R. L. Voute	CGI, TU Delft
	Dr. A.R. Amiri-Simkooei	TU Delft
	F. Dahle	TU Delft
Project Duration:	September, 2023 - May, 2024	

TU Delft

CGI

Abstract

This study explores the application of advanced photogrammetry techniques to enhance the accuracy of Digital Elevation Models (DEMs) derived from satellite imagery. It focuses on refining and integrating the photogrammetry pipeline to accommodate diverse satellite sources effectively. With recent advancements in photogrammetry and the expanding accessibility of satellite imagery, there is a growing opportunity for precise earth surface modeling. This research adapts a traditional photogrammetry pipeline to include state-of-the-art computer vision algorithms, specifically the DISK algorithm for feature detection and LightGlue for feature matching. These enhancements are complemented by tailored adjustments to camera models and projection matrices to suit the unique characteristics of satellite data.

The methodology emphasizes the modification of existing pipelines to optimize the handling of satellite images, incorporating sophisticated feature detection and matching technologies. The performance of these adaptations is rigorously evaluated through extensive analysis using satellite imagery across varied resolutions and environmental conditions. Results from the study indicate marked improvements in the fidelity and accuracy of the generated DEMs, which are substantiated by validation against high-resolution LiDAR ground truth data.

The refined pipeline effectively manages multi-source satellite images and produces terrain models of significantly higher quality, vital for robust geospatial analysis. This work not only bridges the gap between remote sensing and computer vision but also lays the groundwork for future research aimed at improving DEM generation from satellite imagery. This study proposes potential transformative practices in geospatial analysis and supports continued progress in fields such as environmental monitoring, urban planning, and disaster management.

Contents

1

Introduction

1.1. Background and Significance

In recent years, The combination of advances in photogrammetry-the science to make measurement and 3D models from photographs-and the expanding accessibility of satellite imagery have caused an important change in the geospatial sector [3, 8, 21]. This combination has opened new opportunities for detailed earth observation and the careful monitoring of our environment. Notably, methodologies such as Structure from Motion (SfM), a photogrammetric technique that uses overlapping images to produce 3D models of a scene, and Multi-View Stereo (MVS), which further refines these models by considering multiple images of the same area from different viewpoints, have seen significant progress within the computer vision community[36, 38]. The application of these methods has led to great advances in generating Digital Elevation Models (DEMs)—a 3D representation of a terrain's surface, enabling a deeper understanding of the earth's surface complexities. Consequently, numerous software applications are accessible that facilitate high-quality reconstruction [36, 10, 18, 27].

This thesis aims to build upon the foundational work laid out by Kai Zhang et al.[43] , who explored the potential of leveraging state of the art vision 3D reconstruction pipelines for satellite imagery. Their work highlight the minor differences between 3D reconstruction approaches in remote sensing and those developed within the computer vision community. Despite the common objective of 3D reconstruction, the paths taken by researchers in these fields have varied, leading to distinct methodologies and outcomes. This research explores the fusion of computer vision pipelines with satellite imagery to transform DEM generation. Recognizing the potential to enhance the speed and accuracy of satellite-based 3D reconstructions, this thesis aims to bridge a part of the gap between the remote sensing and computer vision communities, leveraging the best of both worlds for improved geospatial analysis.

1.2. Problem Statement

The accuracy of Digital Elevation Models (DEMs) is critical across various applications, including urban planning, environmental monitoring, and disaster management. However, achieving high precision in DEM generation is challenging due to the difficulty in selecting optimal satellite images that are free from atmospheric distortions such as fog and clouds. Current methods often fall short in addressing these issues effectively, leading to inaccuracies in DEMs.

To address these challenges, this research focuses on adapting and enhancing a vision-based reconstruction pipeline for satellite imagery. "Vision-based" refers to the application of computer vision techniques, which use algorithms and artificial intelligence to process and analyze visual data from satellite images. These techniques can detect, interpret, and correct distortions or noise caused by atmospheric conditions, enabling more accurate and reliable extraction of surface information.

The aim of this research is to develop a robust pipeline capable of handling images from multiple satellite sources and operating effectively under diverse environmental conditions. By leveraging vision-based methods, the goal is to ensure that DEMs are generated with high accuracy, even when the source

images are affected by atmospheric distortions. This approach aims for a functional pipeline capable of handling images from multiple satellite sources and under diverse environmental conditions.

The accuracy of DEMs is critical across various applications, including urban planning, environmental monitoring, and disaster management. Despite advancements, achieving high precision in DEM generation faces challenges, notably in selecting optimal satellite images free from atmospheric distortions such as fog and clouds. In an effort to tackle these obstacles, this research adapts and enhances a vision-based reconstruction pipeline for satellite imagery, aiming for a functional pipeline capable of handling images from multiple satellite sources and under diverse environmental conditions.

1.3. Research Objectives

The primary objective of this thesis is to critically evaluate and refine an existing pipeline for DEM creation from satellite imagery, ensuring its adaptability to different satellite sources and improving the accuracy of feature detection algorithms.

The key goals of this thesis include:

- **DTM Derivation from Satellite Imagery:** To understand and establish a systematic process for generating DTMs from satellite imagery, ensuring their quality and reliability for multiple applications. This is done by assessing and understanding the original pipeline.
- **Incorporation of Ground Control Points (GCPs):** To investigate the role and effectiveness of GCPs in refining the DTM generation process, improving the alignment between the DEM and the lidar dataset.
- **DTM Quality Evaluation:** To formulate a comprehensive framework that can assess the quality of the DTMs derived from satellite imagery. This framework should be robust enough to guarantee the DTMs is applicability in diverse scenarios and settings.
- **Improvement and Optimization:** To investigate and pinpoint strategies or modifications that can elevate the quality and reliability of the generated DTMs. This includes delving into every stage of the DTM creation process, addressing potential shortcomings and areas of enhancement.

1.4. Research Questions

The central research question guiding this study is:

How can satellite imagery be used to derive high-quality and reliable Digital Terrain Models for diverse applications?

Subsequently, the research is delineated by the following sub-questions:

- **Relevance of Satellite Image Derived DTMs:** Why are DTMs derived from satellite imagery relevant in today's context and what are the potential benefits?
- **Data availability:** Which data is available and what requirements have to be met for the satellite image data.
- **Effective DTM Generation from Satellite Imagery:** What is an efficient workflow for generating a trustworthy DTM using satellite imagery?
- **Ground Control Points:** How do Ground Control Points and satellite imagery interact to enhance DTM quality?
- **Evaluation of DTM Quality:** How can the derived DTM's quality be validated against known benchmarks or ground truths?
- **Optimization and Enhancement Strategies:** What aspects of the DTM derivation process are most susceptible to quality issues or errors?

1.5. Structure of the Thesis

The following part of this thesis as follows: In Chapter 2, discusses the essential background knowledge that is necessary for a in-depth understanding. In chapter 3, the 3D reconstruction methodology is

explained, including the modification of the pipeline, the integration of a new feature detection algorithm, and the approach to selecting and assessing satellite images with GCPs. Chapter 4 present the Data that is used to for the reconstruction and evaluation. Next in chapter 5 are presents the results and discussion, comparing the performance of the enhanced pipeline to the original and evaluating the quality of the generated DEMs using areal LiDAR data as a benchmark. In the final section, Conclusion and Future Work, Chapter 6 provides a summary of the findings, discusses their implications, and proposes directions for future research in satellite-based DEM generation.

2

Background

The first section (2.1) will provide the background details necessary to build the methods proposed in the next chapter. The first section will cover what DEMs are and why they are crucial in various applications like geology, hydrology, urban planning, and climate change studies. The second section will introduce the role of satellite imagery in generating DEMs, including the types of satellites used. Next (2.3), Photogrammetry is defined and its role in creating DEMs from satellite images.

2.1. Digital Elevation Models (DEMs)

A DEM is simply a model of the “elevation”, which is the height above or below a certain reference point.

A DEM is either a Digital Terrain Model (DSM) or a Digital Surface Model (DTM). A DTM is when the surface of the Earth is the bare-earth, that is no man-made objects or vegetation. Ad DSM is when the surface includes all (man-made) objects and structures on the terrain. The difference can be shown in the figure 2.1. As satellite images include all object and structures on the terrain, when referring to DEM in this thesis, this refers to a DSM. If for other research a DTM is required, it, for example can be retrieved by Ground filtering with Triangulated irregular network (TIN) refinement or by making use of a Cloth simulation filter (CSF) method [22]

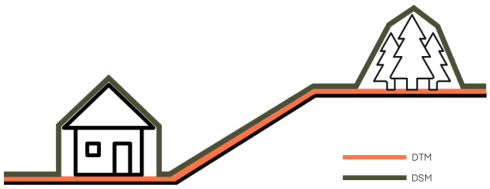


Figure 2.1: Differenct between a DTM (orange) and DTM (grey) [14]

The term 3D in a DSM context can be misleading, as it can refer to three different concepts: 2.5D, 2.75D, and 3D (see figure 2.2). The 2.5D is the most common used version where each location (x,y) in the horizontal plane is assigned to one and only one height z. This thesis wont go to deep in creating dems itself, but the creation of a point cloud with the highest possible quality. A point cloud represent a set of samples that were collected to study the terrain. This thesis uses photogrammetry to gather the samples.

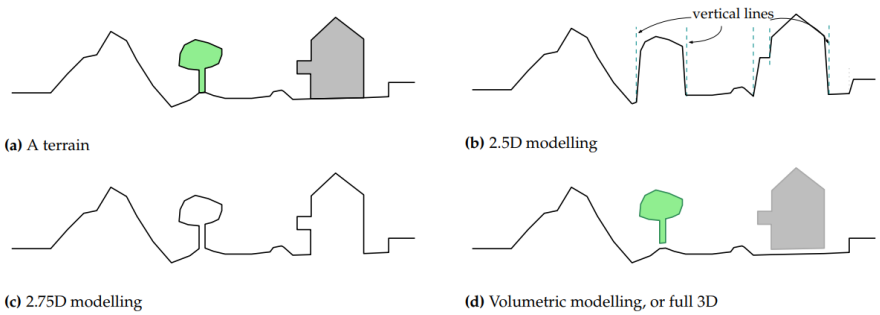


Figure 2.2: Different interpretations of ”3D GIS” as it regards to terrains[14]

2.2. Global Digital Elevation Models (gDEMs)

A global digital elevation model (gDEM) covers extensive geographic areas, typically the entire Earth, which differentiates it from more localized elevation datasets. Because of the enormous scale and prohibitive cost of using airborne or ground surveys, gDEMs rely exclusively on space-borne instruments [22].

Unlike localized surveys, gDEMs must employ sensors that can capture data from orbit. Although some satellites have orbits that do not cover every global latitude, their datasets are still considered global due to their extensive coverage area. These models are crucial for a variety of applications, including environmental studies, geological surveys, hydrological modeling, and disaster management [22]. The acquisition of gDEMs is facilitated through several types of space-borne instruments:

1. Photogrammetry from optical satellite images
2. Interferometric Synthetic Aperture Radar (InSAR)
3. Lidar

While gDEMs typically reflect the Earth’s surface features including vegetation and buildings, making them digital surface models (DSMs), the differences in the data collection methods significantly affect their accuracy and application.

2.2.1. Notable gDEM Initiatives

Several satellite missions have been pivotal in providing comprehensive elevation data:

- **WorldView Constellation (Maxar Technologies, 2007-present):** Utilizes dual images captured during the same orbital pass to derive elevation through photogrammetry. The high spatial resolution (0.31 to 0.50 meters) and short revisit times (1-2 days) make this system particularly effective for dynamic environmental monitoring [**mashaly2018**].
- **Pleiades (CNES, operated by Airbus, 2011-present):** Employs tri-stereo pairs to enhance the accuracy of DSM and DTM extraction, offering high-resolution stereoscopic coverage that is particularly advantageous over complex terrains [**nasir2015**].
- **ASTER (Advanced Spaceborne Thermal Emission and Reflectance Radiometer, 1999-present):** Features two telescopes for capturing stereoscopic images, which helps in reducing mass resource while maintaining a base-to-height ratio conducive to accurate elevation mapping [**cheng2003**].

An analysis of datasets hosted by OpenTopography reveals that local elevation datasets are predominantly available for developed countries. Figure 2.3 illustrates the coverage limitations and the regional concentration of available elevation data.

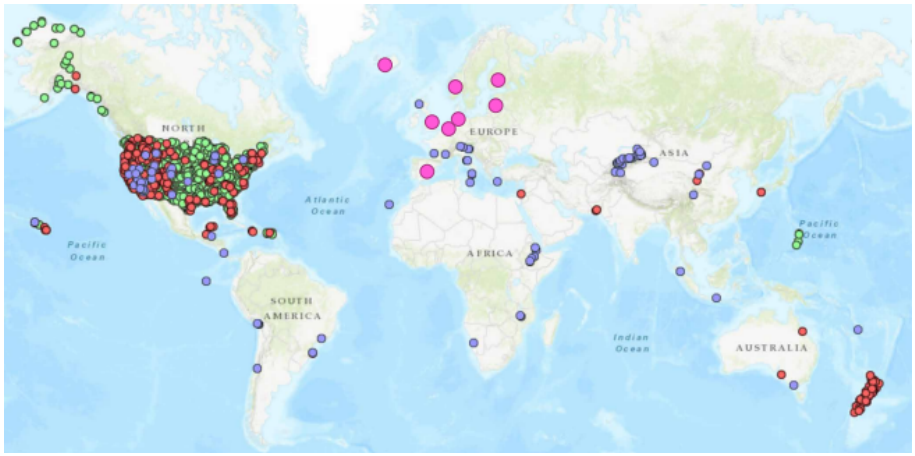


Figure 2.3: OpenTopography coverage with the addition of some European datasets in pink; while there are additional European datasets, no global registry exists for them.[22]

2.2.2. Selection of Satellite Data for This Research

This thesis will focus on the use of optical satellite data for elevation modeling. The next section (2.3) discusses the specific photogrammetry processes employed in this research.

2.3. Photogrammetry fundamentals

Photogrammetry is the science and technology of obtaining reliable information about physical objects and the environment through the process of recording, measuring, and interpreting photographic images and patterns of electromagnetic radiant imagery [9]. The fundamental principle behind photogrammetry is triangulation, which involves capturing photographs of an object or area from multiple angles to determine precise 3D coordinates of points within the object space. This technique has evolved significantly from its early days, driven by advances in photography, computing, and image processing technologies. The process of photogrammetry involves several key steps, each crucial for ensuring accurate 3D reconstruction. With advancements in computer vision, techniques such as Structure from Motion (SfM) and Multi-View Stereo (MVS) have become integral to modern photogrammetry.

2.3.1. Camera Models and Calibration

Photogrammetry relies heavily on precise camera modeling to accurately interpret and reconstruct three-dimensional space from two-dimensional images. The fundamental model used in most photogrammetric applications is the **pinhole camera model**, which serves as an idealized representation of how cameras capture light to form images.

Pinhole Camera Model: The pinhole camera model simplifies the optics of a camera to a single point that captures light. Here, light passes through a small aperture (the pinhole) and projects an inverted image on the opposite side of the camera, typically where the sensor or film is located.

The geometry of the pinhole camera is described mathematically as follows:

- Let $P = (x, y, z)$ be a point in a three-dimensional world coordinate system.
- The corresponding image point $P' = (x', y')$ on the two-dimensional image plane is given by the projection:

$$P' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} f\frac{x}{z} \\ f\frac{y}{z} \end{bmatrix} \tag{2.1}$$

  where f is the focal length of the camera, the distance from the pinhole to the image plane.

Lens distortion is an additional characteristic of a camera lens, in addition to its focal point. With an ideal pinhole camera, each point in the actual world is projected in a straight line onto the camera sensor. However, this is not the case for actual cameras due to optical and manufacturing imperfections. The most common distortions include barrel and pincushion distortions as can be seen in figure 2.4. These are typically corrected in the image processing stage using calibration parameters.

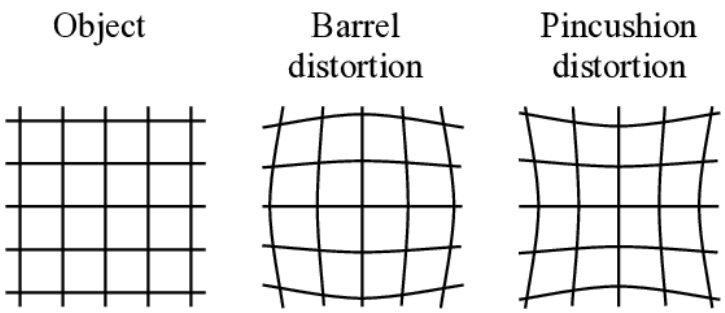


Figure 2.4: Radial distortion, barrel and pincushion. Retrieved from [33]

Camera calibration is a crucial step in photogrammetry that involves estimating the intrinsic and extrinsic parameters of a camera. These parameters define the transformation from the 3D world to the 2D image plane, enabling accurate 3D reconstructions from photographic data. Intrinsic parameters are inherent to the camera's design, and extrinsic parameters describe the camera's position and orientation in space.

Intrinsic parameters include focal length f, the principal point (c_x, c_y), and the skew coefficient which defines the angle between the x and y pixel axes. These are represented in the camera matrix K:

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2.2}$$

where f_x and f_y are the focal lengths along the x and y axes, and s is the skew coefficient.

Extrinsic parameters involve the rotation and translation of the camera relative to a known world coordinate system. They are essential for positioning the camera in space and are given by the rotation matrix R and the translation vector t:

$$\begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \tag{2.3}$$

These parameters are critical for determining the viewpoint of the camera. Given a set of 3D points P_i in the world and their corresponding 2D points p_i in the image, the relationship can be expressed as:

$$p_i = K \begin{bmatrix} R & t \end{bmatrix} P_i = M P_i \tag{2.4}$$

where M is the projection matrix, a combination of intrinsic and extrinsic parameters.

Next the process of camera calibration involves the estimation of the extrinsic and extrinsic parameters of this projection matrix. This process generally involves capturing a series of images from various viewpoints and using these images to infer the camera settings that most accurately map to the observed spatial geometry.

The calibration process identifies correspondences between known points in the world and their projections on the camera's imaging sensor. By analyzing these correspondences, a mathematical relationship is established that models how the camera captures the 3D world.

In solving the camera calibration problem, a linear algebraic approach is typically adopted. This involves generating a system of equations based on the observed correspondences. The objective is to determine the camera parameters that minimize the error between the points projected by the camera model and the actual points observed in the images.

The process often includes solving a minimization problem where the goal is to minimize the overall error between the projected points and the observed image points. This minimization typically employs

optimization techniques that are robust to variations in measurement and can handle large datasets, thereby ensuring a reliable set of camera parameters.

Once the parameters are determined, they provide a comprehensive description of both the camera's internal features, such as lens distortion and focal length (intrinsic parameters), and its position and orientation in space (extrinsic parameters). These parameters are crucial as they affect how the camera will perceive and capture the world, and thus, they are critical for accurate photogrammetric reconstructions.

This calibrated model can then be applied to new sets of image data to predict and map the three-dimensional coordinates of newly observed points, maintaining consistency and precision across various applications in photogrammetry.

2.3.2. Practical Guidelines for Image Acquisition

For effective photogrammetry, adhering to specific guidelines during image acquisition is crucial:

- **Capture Images with Good Texture**: Ensure that the images have sufficient texture to allow for accurate feature detection and matching. Avoid texture-less areas such as clear skies or plain surfaces.
- **Capture Images at Similar Illumination Conditions**: Maintain consistent lighting across images to reduce mismatches. Avoid high dynamic range scenes and specular reflections.
- **Capture Images with High Visual Overlap**: High overlap (60-80%) between images ensures that each feature is visible in multiple images, which is crucial for accurate 3D reconstruction.
- **Capture Images from Different Viewpoints**: Move the camera to different positions to capture diverse perspectives of the scene. This helps in capturing the full 3D structure and reduces errors due to occlusions.

By following these guidelines and understanding the underlying camera models, photogrammetry practitioners can achieve high-quality and accurate 3D reconstructions from photographic data.

2.4. Structure from Motion (SfM)

Structure from Motion (SfM) is a photogrammetric method that reconstructs three-dimensional structures from sequential two-dimensional images. This technology plays a crucial role in applications such as heritage preservation, environmental monitoring, and urban planning, where accurate 3D models are essential. This chapter explores the SfM process, focusing on its implementation and challenges when applied to satellite imagery, which directly addresses a significant component of this research.

2.4.1. The General SfM Problem

SfM involves reconstructing a scene's 3D structure and the motion trajectory of cameras from a series of overlapping images. This complex task includes several critical phases: image acquisition for collecting strategically overlapped images, feature extraction and matching to identify unique points across images, camera pose estimation and point cloud triangulation to determine camera positions and orientations, and bundle adjustment to refine camera parameters and point coordinates to minimize reprojection errors. As shown in Figure 2.5, the SfM processes begin with the collection of images and move through stages of feature extraction, matching, camera pose estimation, bundle adjustment, and ultimately the creation of a dense point cloud.

2.4.2. Image Acquisition

Successful SfM begins with an effective image acquisition strategy, tailored to the project's specifics. This includes planning flight paths for UAVs to ensure comprehensive coverage, choosing the appropriate photo orientation—nadir for direct downward shots or oblique for angled views—to capture the necessary spatial information, ensuring sufficient overlap between photos to facilitate feature matching, and determining the ground sampling distance (GSD), which defines the spatial resolution of the 3D model and is influenced by the altitude and the camera's specifications. Although Ground Control Points (GCPs) are typically used to anchor and scale the reconstructed model accurately, they are not utilized directly in our satellite-based approach due to limitations in the current processing software.

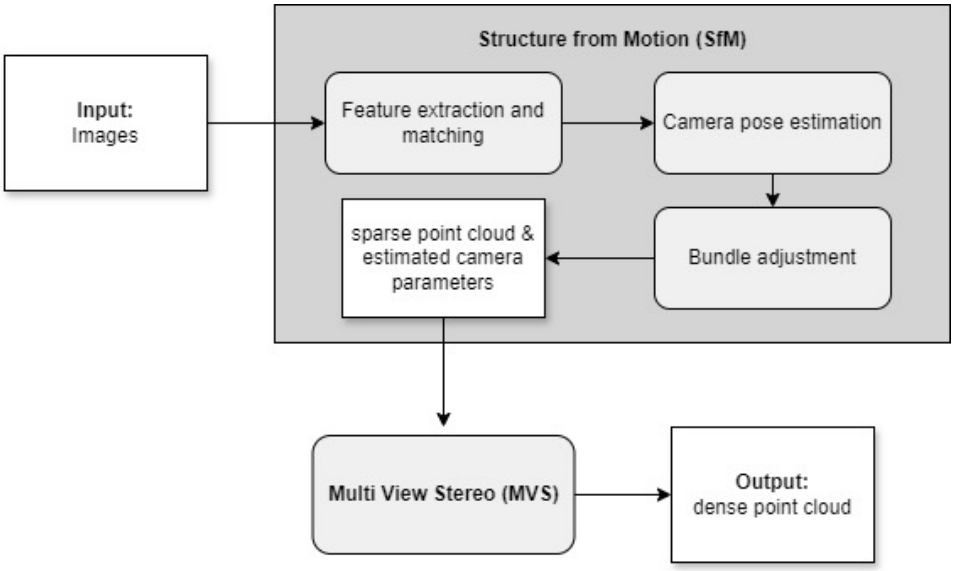


Figure 2.5: Flowchart illustrating the steps in the Structure from Motion (SfM) and Multi-View Stereo (MVS) processes within a modern photogrammetry pipeline.

2.4.3. Feature Extraction and Matching

The next phase involves detecting and describing features that are robust to changes in scale, orientation, and lighting. Algorithms like SIFT and SURF are employed to detect salient features invariant to these changes and then describe them to allow for significant levels of distortion. These features are then matched across multiple images to ensure consistency. Techniques such as RANSAC are used to eliminate erroneous matches and improve the fidelity of the model.

2.4.4. Bundle Adjustment

Bundle adjustment represents the culmination of the SfM process, where both camera parameters and 3D point coordinates are optimized simultaneously. This step minimizes reprojection errors, which are discrepancies between the observed positions of features in the images and their projected positions from the estimated camera parameters and 3D points. It typically involves iterative optimization techniques like the Levenberg-Marquardt algorithm, which refines the camera settings and point estimates to ensure that the projections of the 3D points align with the features detected in the images.

SfM is a powerful tool that translates unstructured 2D image data into structured 3D models, extracting detailed geometric information from a standard set of images. As computational resources and algorithms continue to evolve, the efficiency and accuracy of SfC processes are expected to improve, broadening their applicability in more complex scenarios. This research extends these methodologies to satellite imagery, exploring the unique challenges and necessary adaptations to apply traditional SfM processes effectively at this scale.

2.5. Multi-View Stereo (MVS)

Multi-View Stereo (MVS) techniques extend the capabilities of Structure from Motion by utilizing multiple images of a scene to reconstruct high-accuracy 3D models. While SfM primarily focuses on determining camera positions and sparse 3D point clouds, MVS uses these initial estimates to generate dense and detailed reconstructions of the scene geometry.

2.5.1. Principles and Image Acquisition

MVS algorithms exploit the redundancy of multiple overlapping images to derive depth information by examining the disparities and parallax between images taken from slightly different viewpoints. The core principle relies on establishing a dense correspondence across multiple images, allowing for the precise triangulation of 3D points.

Similar to SfM, the image acquisition process is critical for the success of MVS. Factors such as camera calibration, overlap between consecutive images, and the diversity of viewing angles significantly impact the quality and completeness of the 3D reconstruction. High overlap ensures that each scene point is visible in multiple images, which enhances the reliability and accuracy of the depth estimates.

2.5.2. Depth Estimation and Dense Point Cloud Generation

MVS involves several depth estimation techniques that transform the problem of finding correspondences across images into a search for the best depth for each pixel.

- **Patch-Based Methods**: These methods attempt to find consistent patches across multiple images. They adjust the size and orientation of the patch to maximise the similarity across all views, often using photometric consistency measures.
- **Volumetric Approaches**: Techniques such as space carving remove voxels from a discretized volume if they are not consistent with any image. The result is a conservative estimate of the visible surfaces, ensuring that the reconstructed volume is photo-consistent with all images.
- **Global Optimisation**: Some MVS methods frame the problem as a global optimisation, where a cost function involving all image pixels is minimised. These methods often produce more complete and smoother surfaces but at the cost of increased computational complexity.

Once depth maps are generated for each image, the next step in MVS is to merge these into a single, coherent 3D point cloud. This process involves aligning the depth maps and merging them to form a unified model. Techniques such as point cloud fusion are employed to handle inconsistencies and occlusions effectively.

2.5.3. Challenges

MVS is not without its challenges, which include occlusion. This occurs when parts of the scene are visible in some images but not in others. Effective MVS algorithms need to detect and handle occlusions to avoid artefacts in the 3D model. The accuracy of MVS depends on the quality and consistency of the input images. Low-textured areas, repetitive patterns, and non-Lambertian surfaces can lead to ambiguous matches and incorrect depth estimations [9]. Lastly, MVS algorithms typically require significant computational resources, particularly when processing large image sets or aiming for high-resolution outputs.

Multi-View Stereo, when combined with Structure from Motion, provides a powerful framework for generating detailed and accurate 3D models from a set of images. Advances in computational power and algorithm design continue to push the boundaries of what is possible with MVS, enabling more complex and detailed scene reconstructions than ever before.

2.6. Feature Matching Techniques

Feature matching is a fundamental step in the pipeline of photogrammetric techniques such as Structure from Motion (SfM) and Multi-View Stereo (MVS). It involves identifying and matching points of interest across multiple images to establish correspondences that are crucial for reconstructing the 3D structure of a scene. Effective feature matching is essential for achieving high accuracy and robustness in 3D models.

2.6.1. Scale-Invariant Feature Transform (SIFT)

SIFT works by detecting and describing local features in images. The process involves several key steps:

1. **Scale-space Extrema Detection**: The first step in SIFT is to identify potential interest points that are invariant to scale and orientation. This is achieved by generating a scale space and using the Difference of Gaussian (DoG) method to identify potential keypoints.
2. **Keypoint Localization**: Once potential keypoints are identified, each is refined to eliminate low-contrast points and edge responses to ensure that keypoints are stable. This involves a detailed model fit to nearby data for location, scale, and ratio of principal curvatures.

3. **Orientation Assignment**: Each keypoint is assigned one or more orientations based on local image gradient directions. This step ensures that the keypoint descriptor is rotation invariant.
4. **Keypoint Descriptor**: The local image gradients are measured at the selected scale in the region around each keypoint. These gradients are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

Matching SIFT features between images involves identifying pairs of keypoints with similar descriptors. Typically, the Euclidean distance between descriptors is used to measure similarity, and matches are selected based on the closest match to each keypoint.

2.6.2. LightGlue: Efficient Local Feature Matching

LightGlue is a deep neural network designed to efficiently match local features across images. It enhances the concepts introduced by SuperGlue [34], providing improvements in memory usage, computational efficiency, and training simplicity. LightGlue is adaptive to the difficulty of the matching task, making it suitable for real-time applications and large-scale mapping projects [23]

LightGlue is designed to match local features between images quickly and accurately. Unlike traditional methods relying solely on local descriptors, LightGlue uses a deep learning approach to jointly match sparse points and reject outliers. This method leverages the power of Transformers to learn robust image matching from large datasets, making it highly effective in various environments, including both indoor and outdoor scenes. LightGlue is optimized to be faster and more memory-efficient than its predecessors, such as SuperGlue. This makes it suitable for real-time applications and large-scale mapping projects. The network adapts to the difficulty of the matching task. For easy image pairs with high visual overlap, LightGlue can match features quickly, while for challenging pairs, it allocates more computational resources to ensure accuracy. LightGlue is designed to be easier to train, requiring fewer computational resources. This makes it accessible to a broader range of practitioners and applications.

LightGlue's architecture consists of a series of layers that process the input features. Each layer includes self- and cross-attention units that update the representations of each point based on its relationship with other points in both images. The network also includes a mechanism to dynamically decide when to stop the inference, based on the confidence of the predicted matches. An overview of the arhitecture and each layer can be seen in figure 2.6.

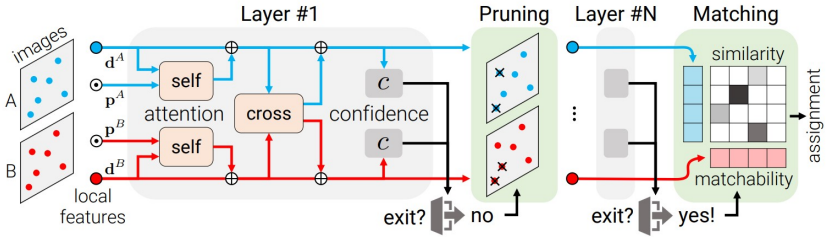


Figure 2.6: The LightGlue architecture, as retrieved from the original paper [23]. Each layer augments visual descriptors with context using self- and cross-attention units, and a confidence classifier determines when to stop inference, pruning confidently unmatchable points and predicting assignments based on pairwise similarity and matchability.

In each attention unit, the state of each local feature is updated by aggregating information from other features. Self-attention units focus on features within the same image, while cross-attention units consider features across both images. This dual attention mechanism allows LightGlue to effectively capture both local and global contexts [23]. Next, LightGlue incorporates an adaptive stopping mechanism that allows the network to halt the inference process early if the predictions are confident. This is achieved by predicting the set of correspondences after each computational block and deciding whether further computation is necessary. This feature significantly reduces the runtime for easy image pairs, making the process more efficient.

For the performance, LightGlue demonstrates to be superior in various benchmarks compared to existing methods [23]. It provides higher accuracy and speed in feature matching tasks, making it a compelling choice for applications like Structure from Motion (SfM) and Multi-View Stereo (MVS).

In the context of photogrammetry, LightGlue offers several advantages:

- **Speed**: Its efficient architecture and adaptive mechanisms enable faster processing of image pairs, which is critical for applications like 3D reconstruction and mapping.
- **Accuracy**: LightGlue's robust matching capabilities ensure high-quality feature correspondences, which are essential for accurate 3D models.
- **Versatility**: It can handle various challenging conditions, such as different lighting and viewpoints, making it suitable for aerial photogrammetry.

2.6.3. DISK Features

DISK (DIScrete Keypoints) is a cutting-edge method in the realm of computer vision, specifically designed for local feature extraction and matching. Traditional methods like SIFT have long been used for these purposes, providing robust performance in various applications, including Structure-from-Motion (SfM) and SLAM. However, DISK introduces a novel approach that leverages Reinforcement Learning (RL) to optimize feature matching in an end-to-end manner, overcoming several limitations of previous techniques [40].

DISK employs a policy gradient method to directly optimize the keypoint detection and matching process. This end-to-end approach ensures that the features extracted are highly relevant to the task of matching, which is a significant improvement over traditional methods that separate these steps. The method utilizes a probabilistic framework to sample keypoints from heatmaps generated by a convolutional neural network (CNN). This allows for a more expressive and flexible selection of keypoints, which can be densely extracted while maintaining discriminative power [40].

By using geometric ground truth to assign rewards to matches, DISK performs gradient descent to maximize the expected reward. This approach ensures that the training process is closely aligned with the actual performance of the feature matching task, leading to better convergence and accuracy. A critical innovation in DISK is its use of analytical expressions for gradients, which mitigates the noise typically associated with Monte Carlo approximations in RL. This enhancement significantly improves the robustness and reliability of the training process [40].

DISK's end-to-end training results in features that yield a higher number of correct matches compared to traditional methods. This increased accuracy is particularly beneficial in applications requiring precise localization and recognition. The probabilistic nature of keypoint selection in DISK allows it to handle variations in scale, rotation, and lighting conditions more effectively than deterministic methods. This robustness is crucial for real-world applications where such variations are common [40]. Despite the complex probabilistic model, DISK maintains efficient training and inference regimes. This efficiency makes it suitable for large-scale and real-time applications, where computational resources and time are critical constraints. By directly optimizing for feature matching, DISK enhances the performance of downstream tasks such as 3D reconstruction and object recognition. This holistic improvement across the pipeline demonstrates the practical advantages of the DISK approach [40].

In conclusion, DISK represents a significant advancement in the field of local feature extraction and matching. Its novel use of RL for end-to-end optimization, combined with a robust probabilistic model and efficient training, sets a new benchmark for performance in computer vision tasks.

3

Methodology

This chapter presents the methodologies adapted and developed to extend a traditional photogrammetry pipeline to satellite imagery applications. It outlines the integration of advanced computer vision techniques into the COLMAP photogrammetry pipeline, ensuring its suitability for handling the complexities of satellite data. Section 3.1 initiates the discussion with necessary adjustments, focusing primarily on the conversion of global coordinate systems to enhance computational stability and precision. Subsequent sections, from 3.3 through 3.4, sequentially explore detailed modifications to the pipeline. These adaptations include the replacement of SIFT feature matching with DISK features and LightGlue matching, as well as the creation of an adapter for SuperView-1 satellite images. Each section provides a logical progression of sophisticated technical adaptations, demonstrating a comprehensive approach to tackling the challenges presented by satellite imagery in photogrammetric applications. The final section, 3.6, evaluates the effectiveness of these modifications, validating the improvements in both accuracy and efficiency of the adapted pipeline.

3.1. Adapting Photogrammetry Pipeline to Satellite Imagery

In recent years, advances in both computer vision and remote sensing have made significant strides in 3D reconstruction techniques [38, 36]. While this computer vision community has largely focused on ground-level imagery, the remote sensing community has advanced methods for satellite-based 3D reconstruction. This chapter describes how a cutting-edge photogrammetry pipeline, called COLMAP, was changed so that it can be used with satellite images. It does this by following the steps that Zhang et al.[43] explained in their work on using vision reconstruction pipelines for satellite images. A flowchart of this adapted photogrammetry pipeline can be seen in figure 3.1.

3.1.1. Tone Mapping

Tone mapping is a critical technique used in photogrammetry to manage the dynamic range of imagery, particularly in adjusting satellite images' brightness and contrast to fit within the display capabilities of standard monitors [24]. This process is essential when dealing with high dynamic range (HDR) content from satellite cameras, which capture a broader spectrum of light intensities than typical display devices can render.

The first step in tone mapping is dynamic range compression. This technique compresses the wide range of light intensities found in satellite images into a narrower, more manageable range suitable for display and analysis. Following this, color correction (or for grayscale images, intensity adjustment) is performed. This step corrects any color distortions introduced by the satellite's sensor or atmospheric interference, ensuring that the displayed images are both visually coherent and accurate.

Contrast enhancement is another crucial component of tone mapping. By adjusting the image contrast, this process emphasizes important textural details and improves the visibility of subtle features, which is particularly beneficial for feature detection and matching in later stages of the photogrammetry pipeline.

Gamma correction is a specific method employed in tone mapping to fine-tune image luminance. The

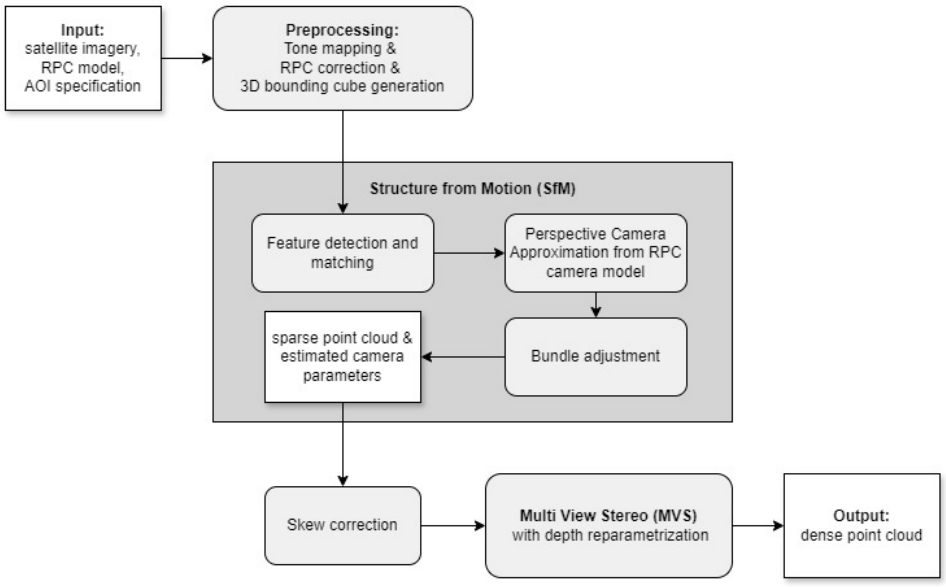


Figure 3.1: Flowchart illustrating the steps in the Structure from Motion (SfM) and Multi-View Stereo (MVS) processes within the adapted photogrammetry pipeline of Zhang et al.

formula for gamma correction is expressed as:

$$I_g = I^{1/2.2}$$

where I_g represents the gamma-corrected image, and I is the original satellite image. This adjustment helps to align the brightness and contrast of the image with human visual perception, enhancing the overall clarity and detail visibility.

Additionally, the tone mapping process involves clipping the intensity values that fall below the 0.5 percentile and above the 99.5 percentile. This filtering step helps to eliminate extreme outliers in the image data, which can skew the visual interpretation and analysis.

Software tools such as MATLAB and Python libraries like OpenCV offer built-in functions for implementing these tone mapping adjustments efficiently. These tools provide robust support for transforming the wide dynamic range of satellite images into a format that is both useful for visual applications and optimized for subsequent photogrammetric processing[43][4].

3.1.2. Rational Polynomial Camera (RPC) Model

The Rational Polynomial Camera (RPC) model is extensively employed in the field of satellite imagery due to its ability to effectively represent complex camera models with a high degree of flexibility [43]. This model is particularly advantageous for applications in remote sensing where detailed, physical models of the camera are unavailable. Most satellite image providers include an RPC model for each dataset, which facilitates accurate geolocation tasks directly from the imagery [12].

The RPC model functions by converting three-dimensional world coordinates (latitude, longitude, and altitude) into two-dimensional image coordinates (u, v). This transformation involves 78 coefficients split between two sets of equations and 10 normalization constants. The model is mathematically expressed as in the following equation 3.1

$$\begin{aligned} u &= \mu_u + \sigma_u \cdot g\left(\frac{x-\mu_x}{\sigma_x}, \frac{y-\mu_y}{\sigma_y}, \frac{z-\mu_z}{\sigma_z}\right) \\ v &= \mu_v + \sigma_v \cdot h\left(\frac{x-\mu_x}{\sigma_x}, \frac{y-\mu_y}{\sigma_y}, \frac{z-\mu_z}{\sigma_z}\right) \end{aligned} \tag{3.1}$$

Here, x, y, z represent the geographic coordinates of latitude, longitude, and altitude, respectively. The functions g and h are ratios of two cubic polynomials, each parameterized by 39 coefficients. These coefficients are derived from the expansion of cubic polynomials, which include terms up to the third power for each variable and their cross-products. Specifically, a cubic polynomial in three variables would normally have 20 coefficients (including a constant term). However, in the RPC model, the constant term in the denominator is typically set to 1 to avoid undefined behaviors during computation, reducing the number of adjustable coefficients in each polynomial to 19. Combined with the 20 coefficients from the numerator polynomial, this accounts for the 39 coefficients for each function g and h.

The normalization constants (μ and σ) scale and shift the input coordinates into a normalized space. This normalization is crucial as it converts large geographical coordinates into a more manageable range, enhancing the numerical stability of the polynomial computations. By adjusting the scale and offset of the coordinates, the RPC model ensures that the polynomial functions operate within their optimal numerical ranges, thus reducing errors and improving the precision of the resulting image coordinate calculations [12].

The versatility of the RPC model lies in its ability to approximate the complex projection geometry of satellite cameras without needing detailed hardware specifications. This feature makes it a preferred choice for generating ortho-photos and extracting Digital Elevation Models (DEMs) from satellite imagery. Its adoption as a standard model across satellite imaging platforms underscores its effectiveness in diverse photogrammetric applications, making it a foundational tool in the geospatial analysis toolkit.

The widespread implementation of the RPC model by satellite image vendors has been driven by its proven success in numerous downstream photogrammetry tasks, such as ortho-photo generation and DEM extraction, as noted in various studies [43][15].

3.1.3. RPC correction

The Earth is approximately ellipsoidal in shape. To pinpoint a location on the Earth's surface, the RPC model employs a global coordinate system composed of latitude, longitude, and altitude. This coordinate system is referenced to a nominal ellipsoid, such as the World Geodetic System 1984 (WGS84). Unlike the Cartesian coordinate system typically assumed by most vision pipelines, the latitude, longitude, and altitude system is not Cartesian. To effectively integrate satellite images into a photogrammetry pipeline, it is essential to convert global geodetic coordinates (latitude, longitude, and altitude) to a local Cartesian coordinate system, specifically the East-North-Up (ENU) system. This conversion enhances numerical stability and simplifies essential calculations in photogrammetric processes [43].

The process for converting coordinates is outlined as follows: To adapt satellite images for use in a photogrammetry pipeline, the process begins by defining a local reference point, which serves as the center for the East-North-Up (ENU) system [44]. Next, the geodetic coordinates (latitude, longitude, and altitude) are transformed into Earth-Centered, Earth-Fixed (ECEF) coordinates, which are Cartesian coordinates centered at Earth's core. Following this, the ECEF coordinates for the reference point are calculated. Differences between the ECEF coordinates of the reference point and those of the point of interest are then determined. Finally, these differences are converted to ENU coordinates using a rotation matrix that aligns the global coordinate system with the local one, facilitating the integration of these coordinates into the photogrammetry workflow.

For practical implementation, MATLAB from MathWorks and the Python library 'pyproj' offer built-in functions to facilitate this conversion[25, 31]. Adopting a local Cartesian coordinate system not only reduces computational complexity and potential for error but also improves the stability of numerical calculations and minimizes distortions commonly associated with global coordinate systems [44]. This ultimately enhances the accuracy of the photogrammetric outputs.

For practical implementation, tools from MathWorks such as MATLAB provide built-in functions (e.g., 'geodetic2enu') to facilitate this conversion [25]. Additionally, Python libraries such as 'pyproj' can also perform these transformations efficiently [31].

Utilizing a local Cartesian coordinate system presents multiple benefits that are crucial for photogram-

metric pipelines, particularly in the integration of satellite imagery. First, lowering the magnitude of coordinate values makes the numbers more stable during calculations. This is especially important for photogrammetry and 3D reconstruction tasks[25]. This aspect of stability is key to ensuring the precision and reliability of measurements and transformations involved in these applications. Additionally, local Cartesian coordinates facilitate the simplification of mathematical operations necessary for camera modeling and bundle adjustment among other photogrammetric processes[6]. This simplification aids in reducing the computational complexity and potential for error in these calculations. Furthermore, by centering the local coordinate system around the specific area of interest, distortions and inaccuracies that are commonly associated with global coordinate systems are significantly minimized, thereby improving the overall accuracy of the photogrammetric output[6].

3.1.4. 3D Bounding Cube Generation

Generating a 3D bounding cube or extracting an Area of Interest (AOI) is essential for focusing the photogrammetry pipeline on a specific region. This process involves defining the AOI by identifying the latitude, longitude, and altitude bounds of the area of interest. Once the AOI is defined, the next step is to calculate the 3D bounding cube that encompasses the AOI. This involves determining the minimum and maximum coordinates in each dimension. Finally, the relevant data within the defined bounding cube is extracted from the satellite imagery. This step reduces computational load and improves processing efficiency by focusing only on the region of interest. This step is critical for managing large datasets typical of satellite imagery, ensuring that computational resources are used efficiently and that the photogrammetry pipeline processes only the necessary data [43][4].

3.1.5. Perspective Camera Approximation

In satellite photogrammetry, the Rational Polynomial Coefficients (RPC) camera model, which often represents complex camera systems, is sometimes approximated by simpler pinhole camera models [43]. This approximation is crucial for integrating satellite imagery into conventional photogrammetric software, enabling the application of advanced structure from motion (SfM) and multi-view stereo (MVS) techniques.

Satellite images frequently utilize intricate camera models like the linear pushbroom model, illustrated in Figure 3.2. This model is characteristic of orbiting satellites where each image row is captured sequentially over time. Given the consistent altitude of satellite orbits, the scene depth—defined as the range of distances from the camera to different points on the Earth's surface visible within a single image frame—can be substantial. However, within this range, the depth parameter Z, which denotes the distance from the camera to a specific scene point, typically varies minimally relative to the average scene depth $\overline{Z}$. This minimal variation in depth across the image allows for the simplification of the complex full perspective model to a more simplified and manageable weak perspective model.

By adopting this simplified model, the computational complexity is reduced, making the processing of satellite imagery more efficient and effective for photogrammetric applications, and enhancing the accuracy and utility of 3D terrain modeling.

Given the minimal depth variation relative to $\overline{Z}$, the perspective projection equations can be simplified as follows:

$$u \approx \frac{f_x}{\overline{Z}} X + \frac{s}{\overline{Z}} Y + c_x, \quad v \approx \frac{f_y}{\overline{Z}} Y + c_y \tag{3.2}$$

where f_x and f_y represent the focal lengths along the x and y axes, respectively, and c_x and c_y are the principal point coordinates. This formula assumes that variations in Z are negligible relative to $\overline{Z}$, which simplifies the projection to closely resemble that of a pinhole camera.

Estimation of the Projection Matrix

To estimate the camera parameters in equation 3.2, the RPC model is sampled to create a series of 3D-2D correspondences. This involves discretizing the expected operational terrain into a grid within the 3D space, projecting each grid point through the RPC model to obtain their corresponding 2D image coordinates. A 3x4 projection matrix P is then computed from these correspondences using the Direct Linear Transformation (DLT) method [13], optimizing projection accuracy by minimizing the error between the projected points and their actual image locations.

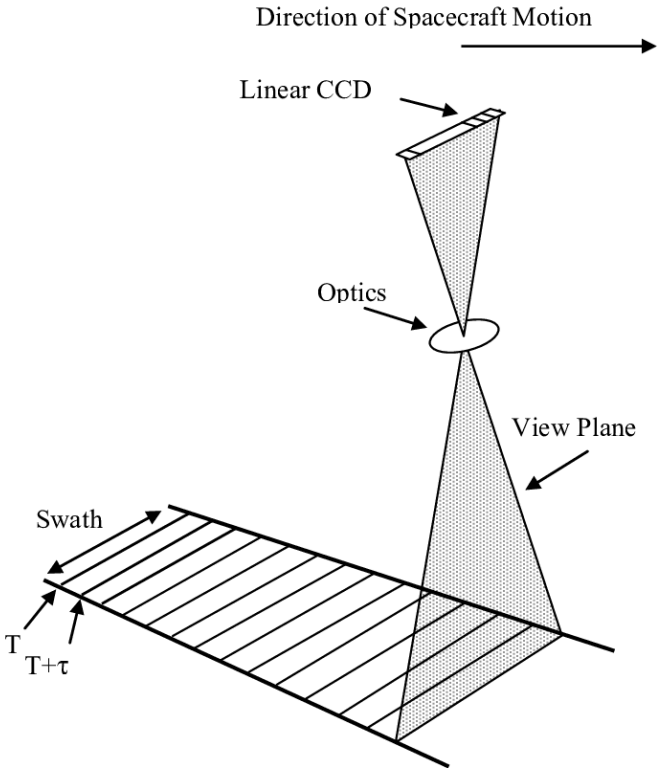


Figure 3.2: Illustration of the push-broom scan technique, where τ represents the integration time. [16]

The advantages of approximating the RPC model to a pinhole camera model include:

- **Simplification of Processing:** It reduces the computational complexity inherent in the RPC model.
- **Enhanced Compatibility:** It allows satellite imagery to be processed using photogrammetric techniques designed for pinhole cameras.
- **Improved Accessibility:** It makes advanced computer vision techniques more accessible for satellite imagery processing.

This approach significantly boosts the practical utility of satellite imagery in photogrammetric applications, promoting more accurate and efficient 3D terrain modeling [12].

3.1.6. Bundle Adjustment

Bundle adjustment is an essential optimization technique used in photogrammetry and computer vision to simultaneously refine camera parameters and 3D point coordinates. It primarily aims to minimize the reprojection errors — discrepancies between observed image points and projections based on estimated 3D points and camera parameters.

In a typical Structure from Motion (SfM) workflow, bundle adjustment enhances the accuracy and consistency of camera models across all images. It involves adjusting both intrinsic parameters (e.g., focal length, principal point) and extrinsic parameters (e.g., orientation and position of the camera), along with the coordinates of sparse 3D points. The goal is to ensure optimal alignment between the projected points and actual image observations across the entire image set.

For satellite imagery processed with a Rational Polynomial Coefficients (RPC) model, bundle adjustment presents unique challenges. The RPC model may exhibit image domain bias drift, affecting geolocation accuracy and object recognition tasks. Additionally, unlike typical SfM pipelines where multiple camera and scene parameters are adjusted, satellite imagery might benefit from only adjusting the principal points (c_x, c_y) to account for any image space translation, while keeping other parameters fixed.

This approach is relevant when images are captured from satellites with highly stable and consistent intrinsic parameters.

To address issues such as projective and gauge ambiguities and to ensure that 3D points remain consistent with geographic locations, a regularization term is often added to the bundle adjustment process. This term penalizes deviations from the original coordinates of the 3D points, ensuring the stability and geographic fidelity of the reconstruction [43].

3.1.7. Skew Correction

In photogrammetry, especially when dealing with satellite imagery, the skew parameter in a camera's intrinsic matrix, which quantifies the non-perpendicularity between the x and y image axes, is often non-zero due to the oblique angle of image capture relative to the nadir. Many computer vision approaches assume zero skew, necessitating a skew correction step to adapt satellite images for integration into standard Multi-View Stereo (MVS) pipelines. This section outlines the method for decomposing the intrinsic camera matrix to remove skew and adapt images for use with conventional photogrammetry software.

The intrinsic camera matrix K incorporating a skew parameter can be expressed and decomposed as follows:

$$\begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{s}{f_y} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_x & 0 & c_x - \frac{s \cdot c_y}{f_y} \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

Here, f_x and f_y represent the focal lengths along the x and y axes, s denotes the skew coefficient, and c_x and c_y are the coordinates of the principal point. The first matrix on the right-hand side accounts for the skew, while the second matrix represents the skew-free intrinsic parameters.

To correct the skew, the inverse of the skew transformation matrix is applied to the images, transforming them into a skew-free reference frame suitable for standard computer vision applications. This correction aligns the images with the zero skew assumption prevalent in most photogrammetric software, enhancing the compatibility and utility of satellite imagery in advanced vision-based analyses.

Figure 3.3 illustrates an example from the TU Delft dataset (Section 4.2) before and after skew correction. In the corrected image, no-data areas on the left widen toward the bottom, showing regions not covered due to the skew correction.

3.1.8. Adapting MVS to Satellite Imagery

Multi-View Stereo (MVS) pipelines traditionally tailored for ground-level images often encounter specific challenges when applied to satellite imagery. This section addresses key issues and introduces adaptations to enable the application of standard MVS approaches to the satellite domain, particularly focusing on depth reparametrization and stable homography computation.

Depth Reparameterization

Considering the challenges with conventional depth due to the significant distance of satellite cameras from Earth—ranging from hundreds to thousands of kilometers—the depth values in satellite imagery exhibit a large mean and comparatively small variance [4]. This scenario poses problems for numerical precision, particularly on consumer GPUs that are limited to single-precision floating points. To address these numerical precision issues that arrise in section 3.1.5, depth is reparameterized using the plane-plus-parallax method [17]. This involves defining a reference plane close to the scene and parallel to the ground plane. The distance from a 3D point to this reference plane is then considered as the reparametrized depth, denoted as m. The mathematical formulation for this reparametrization includes adding a fourth row to the standard 3x4 projection matrix, leading to the following transformation:

$$\begin{bmatrix} u \\ v \\ 1 \\ m \end{bmatrix} = \begin{bmatrix} uZ \\ vZ \\ Z \\ mZ \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ 0 & 0 & \overline{Z} & -\overline{Z}d \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3.4}$$

(a) Before skew correction **(b)** After skew correction

Figure 3.3: Example satellite image of TU Delft before (left) and after (right) skew correction.

Here, m is defined as the reparametrized depth relative to the chosen reference plane. The term $\overline{Z}$ represents the average conventional depth of the scene points calculated during the Structure from Motion (SfM) phase, and d is the offset from the reference plane. This offset ensures that the reparametrized depth values are bounded and more manageable in scale.

3.1.9. Handling the Modified Projection Matrix

Adapting Multi-View Stereo (MVS) pipelines for satellite imagery necessitates significant modifications to how depth information is processed. Unlike traditional MVS techniques, which apply straightforward depth measures derived from pixel displacement in ground-level imagery, satellite imagery requires employing a modified projection matrix. This adaptation leads to several key changes:

- **Depth Computation Adjustments**: Depth computation in satellite-based MVS pipelines involves the introduction of a reparametrized depth variable, denoted as m. This reparametrization recalibrates depth calculations relative to a scene's average depth from a predetermined reference plane, effectively compressing the range of depth values. This adjustment aids in minimizing potential floating-point precision errors, especially on GPUs, and ensures efficient processing across different computational platforms.
- **Enhanced Stability in Depth Processing**: By constraining depth values to a narrower range, the stability of depth-related computations is significantly enhanced. This feature is crucial when using consumer-grade hardware, which may not have the necessary precision for handling large-scale floating-point operations required for extensive 3D reconstructions.
- **Integration into MVS Algorithms**: To ensure that the modified projection matrix seamlessly integrates into existing MVS frameworks, careful calibration is necessary. This involves adjusting the algorithms to accurately interpret the newly measured depth, possibly requiring additional calibration steps or algorithmic modifications.

These modifications improve the accuracy and reliability of 3D modeling from satellite images, overcoming some of the computational limitations associated with consumer-grade hardware. By enhancing the depth parameterization and adapting the projection matrix, photogrammetric techniques can achieve precise results, making advanced 3D reconstruction techniques more accessible and practical across various applications [43].

3.1.10. Handling the Modified Projection Matrix

The adaptation of Multi-View Stereo (MVS) pipelines for satellite imagery introduces significant modifications to the way depth information is processed. Traditional MVS techniques, typically applied to ground-level imagery, treat depth as a straightforward measure derived from pixel displacement. In contrast, the application of a modified projection matrix (equaiont 3.4 in satellite imagery necessitates a different approach:

Depth Computation Adjustments: The introduction of reparametrized depth, denoted as m, shifts the basis of depth calculations. This reparametrization takes into account the scene's average depth relative to a predetermined reference plane, effectively compressing the depth value range. This adjustment not only facilitates a more efficient processing by minimizing the potential for floating-point precision errors on GPUs but also enhances the handling of depth data across various computational platforms.

Enhanced Stability in Depth Processing: Constraining depth values to a reduced range significantly bolsters the stability of depth-related computations. This is particularly beneficial when utilizing consumer-grade hardware, which may lack the precision required for extensive floating-point calculations typically necessary for large-scale 3D reconstructions.

Integration into MVS Algorithms: Incorporating the modified projection matrix within existing MVS frameworks requires careful calibration to maintain consistency and accuracy in depth estimations. Making sure that the MVS algorithms correctly understand the re-measured depth might need extra steps for calibration or changes to the algorithms that are made to fit these changes.

By making these smart changes to the depth parameterization and projection matrix, 3D modelling from satellite images is now much more accurate and reliable. This approach allows photogrammetric techniques to achieve precise results even when faced with the computational limitations of consumer-grade technology, making advanced 3D reconstruction more accessible and feasible in a wider range of settings [43].

3.2. Modification of Feature Detection and Matching

Traditional photogrammetry pipelines often face challenges with feature detection and matching in high-resolution satellite imagery, primarily due to the vast volumes of data and the complexity of image features. The photogrammetry pipeline proposed by Zhang et al. [43] traditionally utilized the SIFT algorithm for feature detection and matching, which is outlined in the top half of Figure 3.4. This section outlines enhancements made to the pipeline, incorporating advanced machine learning techniques alongside DISK features [40] and LightGlue matching [23]. The steps of the adaption are outlined in the bottom half of figure 3.4.

3.2.1. Feature extraction with DISK

The feature extraction process is a pivotal component in the photogrammetry pipeline, especially for applications involving high-resolution satellite imagery where feature density and computational efficiency are of significant concern. To meet these requirements, the DISK (Deep Iterative Subspace Keying) [40] feature extractor was chosen because it can work with large datasets easily and reliably find and describe features.

The DISK feature extractor, implemented using the PyTorch library, was initialized to detect up to 3000 features per image. This parameter was set based on a balance between computational limits and the expected density of features in satellite images, which typically present complex scenes with high detail. The choice of 3000 features is the same as was used for SIFT feature matching and makes it easier to do a qualitative comparison between these different methods [43].

DISK feature extraction works by using a convolutional neural network (CNN) architecture that has been trained to find and describe image features that do not change when the size, rotation, or lighting changes. This method is particularly suited for satellite imagery, which can vary widely in appearance due to differing capture conditions [40].

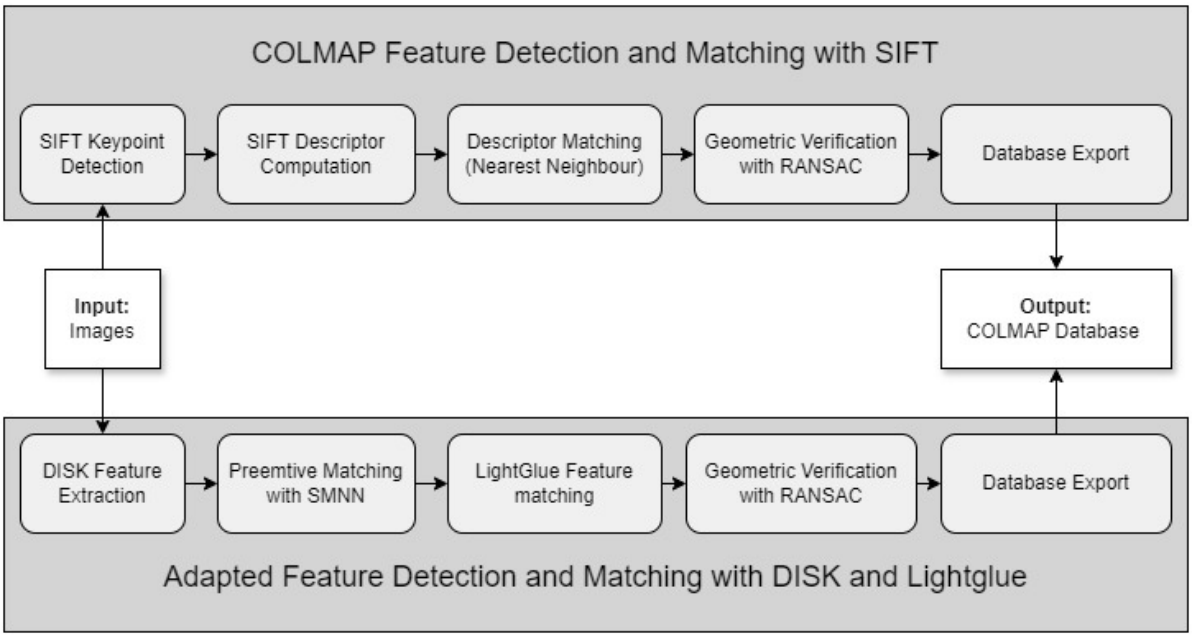


Figure 3.4: Comparison of traditional COLMAP feature detection and matching with SIFT and the adapted method using DISK and LightGlue.

Image Loading and Preprocessing

Each satellite image was loaded using the OpenCV library, which provides comprehensive support for image processing tasks. This step was crucial for handling various image file formats and data types present in satellite image datasets. Post-loading, images were converted to a consistent RGB color format using OpenCV functions. This standardization is essential as satellite images often arrive in diverse spectral bands and color formats depending on the sensor. Normalizing to RGB simplifies subsequent processing steps and ensures consistency in input data for the DISK algorithm [40].

Feature Detection with DISK

The process begins with the application of the DISK algorithm to each preprocessed satellite image to detect stable and distinct points, referred to as features.DISK, utilizing advanced machine learning techniques, scans the images to identify points that demonstrate consistency and distinctiveness under various imaging conditions. This detection process leverages the computational power of deep learning platforms like PyTorch, which facilitates the rapid processing of large datasets typically associated with high-resolution imagery. Once features are detected, DISK generates a unique descriptor for each one. These descriptors capture the essential local characteristics of each feature in a robust, compact format that is highly suitable for subsequent matching processes. The descriptor provides a comprehensive representation of the feature, allowing it to be effectively recognized and matched across different images, even under varying lighting and angle conditions. The strength of these descriptors lies in their ability to support precise and reliable comparisons, which is crucial for accurate image matching and 3D modeling.

3.2.2. Preemptive Matching

Preemptive matching is an essential step in the photogrammetry pipeline designed to enhance computational efficiency during feature matching. This process involves selectively narrowing down potential feature matches before proceeding to the more computationally demanding detailed matching phase. By reducing the number of feature comparisons early in the process, preemptive matching helps to accelerate the overall workflow and minimize computational resource usage, making the pipeline more scalable and efficient[41].

3.2.3. Preemptive Matching in Photogrammetry

Preemptive matching is a crucial optimization step in the photogrammetry pipeline tailored to enhance the efficiency of feature matching. By filtering potential matches between images at an early stage, this process significantly reduces the computational load during the detailed matching phase. This section details the approach used to implement preemptive matching, focusing on its role in streamlining the matching process for high-resolution satellite imagery.

Overview of Preemptive Matching

In the adapted feature detection and matching pipeline, preemptive matching serves as the first level of filtration. Its primary purpose is to quickly narrow down the vast number of potential matches based on preliminary criteria, thus ensuring that only the most likely matches are forwarded to the more computationally intensive stages of geometric verification and detailed matching.

Operational Mechanism

The preemptive matching process begins by examining all possible pairs of images within the dataset. For each pair, features previously extracted and stored are retrieved, and a subset of these features—typically up to a predefined number like 3000 features per image—is considered for initial matching. This limitation is strategically set to balance between thoroughness and computational feasibility.

The matching itself is conducted using the Shared Nearest Neighbors (SMNN) approach, which evaluates the similarity between feature descriptors across the two images. SMNN is particularly suited for this task as it effectively identifies potential matches by considering not only the direct similarity between features but also the consistency of their mutual nearest neighbors, thus enhancing the robustness of the matches.

Match Selection and Storage

The outcome of this preliminary matching stage is a list of image pairs along with the count of matches identified between them. These results are then stored systematically, typically in a structured format like a Pandas DataFrame, which is serialized into a file for persistence. This data serves as the input for subsequent processing stages, where detailed analysis and verification of these preliminary matches are performed.

By implementing preemptive matching, the photogrammetry pipeline efficiently manages the extensive data involved in processing high-resolution satellite images. This step not only speeds up the overall feature matching process but also ensures that computational resources are concentrated on the most promising matches, thereby enhancing the efficiency and accuracy of the final photogrammetric products.

3.2.4. Feature Matching Using LightGlue

LightGlue plays a pivotal role in the feature matching stage of the photogrammetry pipeline, particularly after the initial preemptive matching phase has narrowed down potential matches. As a sophisticated algorithm, LightGlue utilizes deep learning techniques to enhance the accuracy and robustness of matches across intricate datasets, like those derived from satellite imagery.

Integration of LightGlue in the Matching Process

Following the initial filtration accomplished by preemptive matching, LightGlue is employed to meticulously analyze and refine these preliminary matches. This is crucial for ensuring that the matches are not only based on feature similarity but are also geometrically consistent across images, thereby enhancing the reliability of the photogrammetric analysis.

Utilization of Descriptors and Local Affine Frames

The process begins with the retrieval of descriptors and Local Affine Frames (LAFs) from the features detected by the DISK algorithm. Descriptors provide a robust representation of the image features, encapsulating critical information about their appearance in a compact form. LAFs contribute additional data about the local geometry of the features, including orientation and scale, which are essential for maintaining consistency across different viewpoints.

Calculation of Matching Scores

LightGlue calculates matching scores by analyzing the descriptors for similarity and examining the geometric consistency provided by LAFs. The similarity of descriptors ensures that the features under comparison are indeed representations of the same point in different images. Meanwhile, the geometric consistency checks align with the physical laws of perspective and projection, ensuring that matches are not only visually similar but also align correctly in three-dimensional space.

Output and Validation of Matches

Each match generated by LightGlue is quantified with a score that reflects the confidence in its accuracy. These scores are crucial as they help to filter out less reliable matches in subsequent processing steps. The output consists of pairs of indices, each pair linking a feature in one image to a feature in another image, based on the highest scores of matching validity.

By employing LightGlue, the photogrammetry pipeline benefits from an advanced level of precision in feature matching, which is critical for generating accurate and reliable 3D models from satellite imagery. The use of both descriptor similarity and geometric consistency ensures that the matches are robust, making LightGlue an indispensable tool in modern photogrammetric workflows.

3.2.5. Match Filtering

Match filtering is an essential step in the photogrammetry pipeline designed to enhance the accuracy and reliability of the 3D reconstruction process. After preliminary matching, even with advanced algorithms like LightGlue, the set of putative matches often contains false positives. These are incorrect matches that appear valid due to similar feature descriptors but do not represent actual correspondences in the scene.

A primary criterion for filtering is geometric consistency, which involves verifying that matches comply with the expected geometric transformations given the camera positions and orientations. This is assessed through a robust estimation technique RANSAC (Random Sample Consensus), which identifies inliers among the matched points that conform to a plausible geometric model.

Matches are also evaluated based on their scores provided by LightGlue, which reflect the confidence in the correctness of each match. A threshold is set to exclude matches with scores below a certain level, indicating lower confidence in their validity. This threshold is carefully chosen based on the distribution of match scores and the specific requirements of the reconstruction task.

3.2.6. Database Export

The export of a well-structured database compatible with COLMAP is a critical step in preparing the photogrammetry pipeline for 3D reconstruction. This process involves organizing and storing the features and filtered matches in a way that aligns with the expectations and requirements of the COLMAP software.

The data prepared and stored includes not just the raw image and feature data but also the relational information that links these elements together, such as which features correspond to which images and how images relate to each other through matched features. This comprehensive structuring is critical for ensuring the data can be effectively utilized by COLMAP's reconstruction algorithms.

For detailed specifications on how the data should be formatted and the schema that should be used within the COLMAP database, the COLMAP database documentation is used[37]. This documentation provides essential guidelines on structuring data in a way that is compatible with COLMAP, ensuring that the software can efficiently process the data for 3D reconstruction.

3.3. Modification of Feature Detection and Matching

Traditional photogrammetry pipelines often face challenges with feature detection and matching in high-resolution satellite imagery, primarily due to the vast volumes of data and the complexity of image features. This section outlines enhancements made to the pipeline, incorporating advanced machine learning techniques alongside DISK features and LightGlue matching. These improvements are specifically tailored to enhance the accuracy and efficiency of feature matching and detection processes. By leveraging these modern computational methods, the pipeline significantly boosts the precision and speed

of 3D reconstructions derived from satellite imagery, facilitating more robust and reliable geospatial analyses.

3.4. Satellite Adapter for SuperView-1 Images

The integration of SuperView-1 (SV1) imagery into the existing photogrammetry pipeline, originally optimized for WorldView-3 (WV3) images, requires adjustments to accommodate the distinct specifications and formats of SV1 data. This section details the implementation of an input adapter designed to preprocess SV1 images to ensure compatibility and maintain processing efficiency with the pipeline.

Unlike WV3 images which are delivered in NITF format, SV1 imagery is provided in Tag Image File Format (TIFF). TIFF files containing SV1 imagery are identified and processed using the Geospatial Data Abstraction Library (GDAL), a robust tool for reading and writing raster data. This step involves extracting both the image data and associated metadata from each file. Special attention is dedicated to the Rational Polynomial Coefficients (RPC) data, crucial for accurate geolocation. Adjustments or transformations are often necessary to align the RPC data from SV1 with the expected format used by the photogrammetry tools in subsequent processing stages.

3.4.1. Image Cropping and RPC Adjustment

Each image is cropped based on predefined Area of Interest (AOI) parameters to ensure that only relevant portions of the images are processed. This optimization reduces computational resources and aligns images spatially with the targeted area for modeling. Post-cropping, RPC parameters are adjusted to reflect the new image dimensions, ensuring accurate geospatial referencing through the subsequent processing stages.

3.4.2. Metadata Adjustment and Storage

Following image cropping, metadata, particularly RPC values, are adjusted to match the new dimensions of the images, ensuring precise geolocation. These adjustments are crucial as they directly influence the accuracy of subsequent geospatial processing. Adjusted metadata, including modified RPC values and new image dimensions, is systematically stored in a structured JSON format. This structured metadata storage is essential for subsequent stages of the pipeline, ensuring that all processing steps have access to accurate and necessary data. The images are also converted to a PNG format, optimizing them for the feature detection and matching processes that follow.

3.4.3. Adaptation Benefits and Challenges

The adaptation of the photogrammetry pipeline to include SV1 satellite imagery expands its applicability to a broader range of data sources, which is particularly beneficial given the free availability of SV1 data in regions like the Netherlands. Moreover, this adaptation introduces challenges such as adjusting for differences in metadata formats and scales between SV1 and WV3 images, which were addressed through specific software tools and procedural adjustments.

This enhanced adaptability of the pipeline not only allows for the processing of a diverse array of satellite imagery but also improves the overall robustness and efficiency of the data processing workflow, ensuring that the pipeline remains effective and relevant in various geospatial analysis applications.

3.5. Digital Surface Model Production from Point Clouds

This section outlines the methodology employed to transform dense point clouds obtained from stereo photogrammetric processes into a 2.5D Digital Surface Model (DSM). The process of transforming georeferenced 3D points into a Digital Surface Model (DSM) consists of several steps:

Point Grid Projection

Initially, the georeferenced 3D points are positioned onto a predefined grid based on their geospatial coordinates. This projection is mathematically represented by the equation:

$$\text{index} = \left\lfloor \frac{\text{coordinate} - \text{offset}}{\text{resolution}} \right\rfloor \tag{3.5}$$

where `offset` is the coordinate of the upper left corner of the grid, and `resolution` is the grid resolution set to 0.5 meters, indicating that each grid cell represents a 0.5 meter by 0.5 meter area on the ground. This method ensures that each point is accurately placed within the grid, corresponding to its geospatial coordinates.

Elevation Interpolation

After positioning, the elevation values are assigned to the corresponding grid cells. If multiple points map to the same grid cell, the cell assumes the elevation of the highest point, which represents the topmost surface visible from above. Assigning the highest elevation value to each grid cell in a Digital Surface Model (DSM) ensures that the model accurately represents the topmost surfaces, like buildings and tree canopies, visible from an aerial view, providing a true depiction of the terrain's most significant features. Grid cells lacking direct elevation data are interpolated using a median filter, smoothing the DSM and filling gaps in data coverage. This is implemented using the `numpy_groupies.aggregate` function [26], which efficiently handles large arrays of indices and values to create the DSM.

DSM Smoothing

The DSM undergoes a median blur filtering process to enhance visual clarity and usability. This step reduces noise and minor irregularities. The medianBlur from OpenCV[29] function computes the median of all the pixels within the kernel area and replaces the central element with this median value. Unlike other filters where the central element could be a newly calculated value or an existing pixel value, median blurring always replaces the central element with a pixel value from the image. This process significantly reduces noise. The kernel size used should be a positive odd integer; in this case, a kernel size of 3x3 pixels is chosen.

The final DSM is outputted in GeoTIFF format for use in Geographic Information Systems (GIS) and other applications. Additionally, a visual representation of the DSM is generated in JPEG format to provide an immediate, understandable view of the terrain..

3.6. DSM Evaluation

Evaluating the accuracy and reliability of the Digital Surface Model (DSM) generated from satellite imagery forms a crucial part of this research. This section details the comprehensive validation approach against high-resolution LiDAR data. As ground truth, lidar-derived topographic data serve as a fundamental benchmark to assess the DSM's fidelity. The evaluation process not only emphasizes the alignment of coordinate systems but also quantifies the DSM's accuracy through statistical metrics, ensuring that the transformations and adaptations implemented in the photogrammetry pipeline deliver precise and usable geospatial data. Subsections discuss the critical steps involved in this process, from RDNAP transformations to vertical co-registration and numerical evaluations, outlining how each contributes to a robust assessment of the DSM.

3.6.1. Ground truth, Lidar

As ground truth is the Actueel Hoogtebestand Nederland (AHN) data used[1]. This is a fundamental component of this research, providing high-resolution lidar-derived topographic data essential for validating the digital elevation models (DEMs) created from satellite imagery.

The AHN, or Actual Height Model of the Netherlands, is a series of nationwide lidar surveys that provide detailed and accurate topographical data of the Dutch landscape. It is periodically updated to reflect ongoing changes in the terrain. The version used in this research, AHN4, represents the latest iteration, offering unprecedented detail and accuracy.

The AHN data is referenced to the Rijksdriehoekstelsel (RD), the Dutch national coordinate system, which uses the Amersfoort coordinate as its origin. This system is aligned with the European Terrestrial Reference System 1989 (ETRS89) for horizontal positioning, ensuring compatibility with international geospatial data standards. Vertically, AHN data is tied to the Normaal Amsterdams Peil (NAP), which serves as the national standard for elevation measurement in the Netherlands. The NAP is equivalent to the mean sea level at Amsterdam, acting as the zero elevation point for height measurements across the country. Because of this it can also be referenced to as RDNAP.

3.6.2. RDNAP Transformation

Changing the coordinate systems to line up the DEMs made from satellite data with the high-resolution AHN lidar data is an important step in making sure that the datasets can be compared and layered correctly. This transformation involves converting the coordinate reference system (CRS) from the global WGS84 used by the satellite imagery to the local Rijksdriehoekscoördinaten (RD or RDNAP) used by the AHN data.

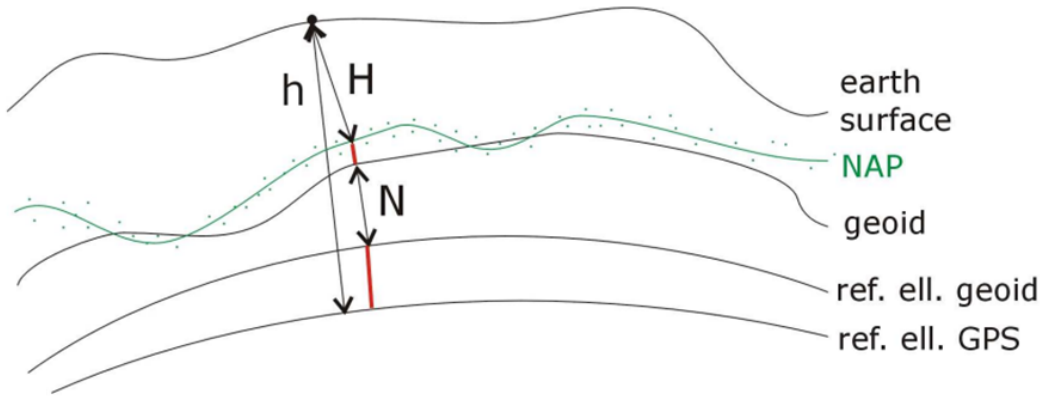


Figure 3.5: Display of the various vertical datums. There may be disparities in a transition between various height reference surfaces since different surfaces are used as height references. Retrieved from [30]

Necessity of RDNAP Transformation

Satellite images used in this thesis are initially referenced in the WGS84 coordinate system, which is a global reference system for latitude, longitude, and altitude. However, the AHN lidar data is referenced to the Dutch national standard, RDNAP, which uses a different projection and vertical datum. Aligning these datasets requires a precise transformation to ensure that all data points from both sources correlate accurately on the same scale and reference.

RDNAP Transformation Process

The RDNAP transformation involves several steps and utilizes the EPSG:7415 transformation, which includes shifting from ETRS89 (aligned closely with WGS84 for Europe) to RDNAP. This transformation includes adjustments for both positional shifts and vertical alignment. Here's how the transformation is applied:

The RDNAP transformation process uses the EPSG:7415 transformation, transitioning from ETRS89, which aligns closely with WGS84 for Europe, to RDNAP. This comprehensive transformation includes adjustments for both positional shifts and vertical alignment. The process begins with coordinate conversion where coordinates in the WGS84 system (latitude, longitude, altitude) from the satellite data are converted to their equivalents in the ETRS89 system. Although this step usually results in negligible positional changes, it's essential for ensuring formal accuracy.

Following conversion, the ETRS89 coordinates are projected into the RD New (RDNAP) coordinate system using a specific projection formula. This step transforms the geographic coordinates into a planar system, which is widely used across the Netherlands for all official spatial data. Lastly, the altitude data, initially aligned with the global mean sea level, undergoes vertical adjustment to align with the Normaal Amsterdams Peil (NAP), the Dutch standard for elevation. This ensures that all height data correspond accurately to local elevation references, making them comparable with national data sets. is the reference level for elevations in the Netherlands.

Implementation in Python

The transformation is implemented using the PyProj library in Python, which provides tools to perform complex projections and transformations seamlessly. The specific pipeline used for the RDNAP transformation involves a series of steps that convert geographic coordinates to grid coordinates and adjust altitudes from the ellipsoidal heights to orthometric heights relative to NAP.

3.6.3. Vertical Co-Registration

Initial comparisons of the digital surface model generated during this research and the corresponding ground-truth LiDAR DEM revealed a noticeable vertical offset. This misalignment likely originated by the camera models used for image capture, or inconsistencies within the processing pipeline. Although horizontal discrepancies were also observed, their impact was minimal and thus not addressed within the scope of this thesis. The primary focus was to correcting the vertical misalignment to enhance the DEM's precision.

Vertical co-registration is the process of aligning two DEMs to ensure their elevation values match as closely as possible [28]. This adjustment is critical for addressing any vertical offsets that may have arisen due to data acquisition errors or inconsistencies within the processing pipeline. The vertical co-registration process included the following steps:

1. **Identification of Corresponding Points:** Points were chosen based on their high likelihood of accurate reconstruction, with a preference for stable, well-defined features such as building rooftops and road surfaces. This identification was performed visually to ensure consistency across both DSMs.
2. **Calculation of Vertical Differences:** The vertical difference between the generated DSM and the ground-truth DSM was calculated at each selected point. This step involved comparing the elevation values from both models at these points to determine the extent of vertical misalignment.
3. **Application of Vertical Adjustments:** An average of these vertical differences was computed and used as the basis for adjusting the entire DEM. This average adjustment was applied uniformly across the generated DEM to correct the vertical positioning.

The implementation of vertical co-registration involved applying the calculated adjustment across the entire DEM, effectively aligning it with the elevation values of the ground-truth LiDAR DSM.

The vertical co-registration process ensures that the elevation values in the generated DEM closely align with those of the ground-truth DEM. Accurately adjusting for vertical errors enhances the integrity and utility of the DSM, providing a reliable basis for further geospatial analyses. This step not only mitigates systematic errors introduced during data collection and model creation but also improves the overall quality of the geospatial outputs.

3.6.4. Numerical Evaluation

Following the vertical co-registration of the two DSMs, it is crucial to evaluate the accuracy to ensure that the alignment is both theoretically sound and practically effective. This section outlines the methodology used to quantify the accuracy of the aligned DSMs using established statistical metrics. The accuracy assessment involves calculating several error metrics, such as the mean error, the root mean square error (RMSE), the standard deviation (SD) of the differences, and the Completeness Value (CV) between the aligned DSM and the ground-truth DSM (gt) for every grid cell.

The mean error (ME) is calculated using the formula:

$$\text{Mean} = \frac{\sum_{i=1}^{n} d_i}{n} \tag{3.6}$$

The standard deviation (SD) of the differences is computed using:

$$\text{SD} = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \mu)^2}{n}} \tag{3.7}$$

The root mean square error (RMSE) is determined by:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} \tag{3.8}$$

where d_i represents the difference between the elevation values of the aligned DSM ($h_i^{aligned}$) and the ground-truth DSM (h_i^{gt}), μ is the mean of these differences, and n is the total number of grid cells or pixels.

Additionally, the Completeness Value (CV) for thresholds $\epsilon_1, \epsilon_2, \ldots, \epsilon_k$ is used to assess the percentage of data points that fall within various predefined error thresholds, offering insight into the completeness of the DSM data. The CV is calculated as follows:

$$\text{CV}(\epsilon_i) = \left(\frac{\text{Number of pixels with } |d_i| \leq \epsilon_i}{\text{Total number of pixels}} \right) \times 100 \tag{3.9}$$

These metrics provide a comprehensive understanding of the quality, ensuring that the vertical co-registration process has effectively minimized the discrepancies between the two DSMs.

4

Data Description

This chapter details the datasets utilized in this thesis, encompassing high-resolution satellite imagery and precise lidar measurements. Each dataset's characteristics, including the acquisition parameters, geographic coverage, and data formats, are outlined. The datasets from Argentina and the Netherlands represent varied urban and semi-urban environments, offering a basis for the assessment of the proposed photogrammetry pipeline.

4.1. Benchmark Dataset of San Fernando, Argentina

Due to the lack of available public datasets, M. Bosh et al. created a public multiple view stereo benchmark dataset in 2016 [39]. This dataset facilitates research advancements in the field by enabling the application and customization of methods to real-world problems. Developed mainly for the IARPA Multi-View Stereo 3D Mapping Challenge, it includes fifty Digital Globe WorldView-3 panchromatic and multispectral images covering a 100 square kilometer area near San Fernando, Argentina. High-resolution airborne lidar ground truth data is also provided for a 20 square kilometer subset of this area, enhancing the dataset's utility for detailed surface modeling.

4.1.1. Source Imagery

The source imagery comprises fifty WorldView-3 panchromatic, visible, and near-infrared (VNIR) images collected between November 2014 and January 2016, and one short-wave infrared (SWIR) image from November 2015, all provided in National Imagery Transmission Format (NITF). For this thesis, only the panchromatic and visible images are utilized. These images cover the area near San Fernando, Argentina, as shown in Figure 4.1. The ground sample distance (GSD) for panchromatic images is approximately 0.31m, and for VNIR images, it is about 1.24m, as documented by WorldView-3 Instruments [7].

4.1.2. Ground Truth Lidar

Airborne lidar data was collected in June 2016, covering approximately 20 square kilometers overlapping with the image coverage. This data supports the production of 30cm gridded Digital Elevation Model (DEM) products, facilitating metric comparison with MVS-generated point clouds [3].

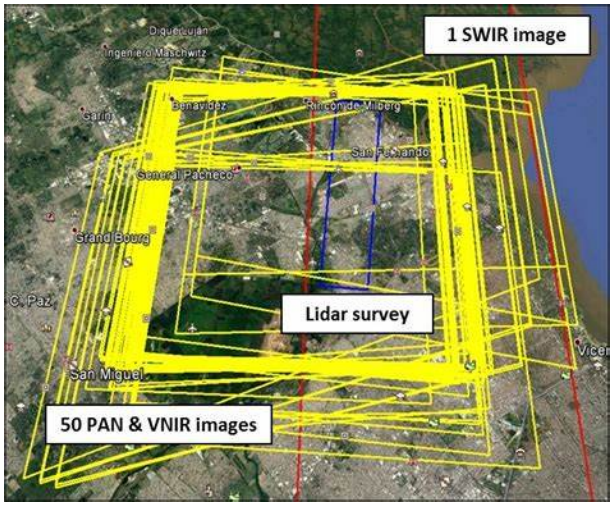


Figure 4.1: Benchmark dataset visualized in Google Earth with polygons for PAN & VNIR images (yellow) and Lidar coverage (blue) Retrieved from [3].

4.1.3. Used Datasets

Due to the extensive size of the source imagery and the limited availability of ground truth data, the source data was cropped into smaller segments. Following the approach of K. Zhang, three sites from the IARPA MVS3DM dataset[39] were selected for detailed analysis in this thesis [43]. These sites were chosen based on their diverse terrain features and data quality, providing a robust basis for evaluating photogrammetric techniques. Figure 4.2 shows examples of cropped images and corresponding lidar data for these sites.

4.1.4. Accessing the IARPA MVS3DM Dataset

It is important to note that the previously referenced location for the IARPA MVS3DM dataset in some documents is incorrect. To facilitate accurate access for research purposes, the dataset can be downloaded using Amazon Web Services (AWS). Researchers interested in utilizing the IARPA MVS3DM dataset should use the following AWS command to copy the data to their local systems:

```
aws s3 cp s3://spacenet-dataset/Hosted-Datasets/MVS_dataset/ \
    /local/path/to/dataset/ --recursive
```

This command will recursively copy all files from the specified S3 bucket to a local directory on your machine. Make sure to replace ‘/local/path/to/dataset/‘ with the actual path where you wish to store the dataset on your local system.

(a) Site 1 - Panchromatic **(b)** Site 1 - Multispectral **(c)** Site 1 - GT Lidar

(d) Site 2 - Panchromatic **(e)** Site 2 - Multispectral **(f)** Site 2 - GT Lidar

(g) Site 3 - Panchromatic **(h)** Site 3 - Multispectral **(i)** Site 3 - GT Lidar

Figure 4.2: A cropped panchromatic and multispectral image of site 1 to 3 with the corresponding Lidar GT

Figure 4.3: Color bar for displaying height map of the Lidar data in figure 4.2

Table 4.1: Overview of the selected data for each site, showing the number of views, average number of pixels, and area.

Site	# Views	Avg. # Pixels	Area (km^2)
Site 1	47	4.22M	0.464
Site 2	47	1.27M	0.138
Site 3	40	1.44M	0.150

4.2. Datasets Delft and The Hague, Netherlands

This section details the datasets from the Netherlands used in this thesis, which are publicly available through Satellietdataportaal.nl [35] and sourced from the SuperView-1 satellite. These datasets are essential due to their free accessibility and high resolution, suitable for advanced photogrammetric analysis.

4.2.1. Source Imagery and Coverage

The datasets cover high-resolution panchromatic imagery from two significant locations in the Netherlands: a part of the TU Delft campus and an area in the city of The Hague. The SuperView-1 satellite's availability of numerous images with the highest resolution led to the selection of these locations. Details regarding the number of views, pixel counts, and area coverage are summarized in Table 4.2 below. This imagery has a ground sample distance (GSD) of 0.5 meters, offering detailed textural information of the urban landscape. Although multispectral imagery with a resolution of 2 meters is also available, it was not utilized in this research. An overview of the satellite coverage and area of interest for Delft can be seen in figure 4.4

Table 4.2: Overview of the selected data for the Netherlands sites, showing the number of views, average number of pixels, and area covered.

Site	# Views	Avg. # Pixels	Area (km^2)
TU Delft	21	1.74M	0.435
Den Haag	21	2.33M	0.582

Figure 4.4: Satellite dataset (yellow) visualized on an aerial orthophoto, illustrating the coverage of the SuperView-1 datasets used. The area of interest is marked in red.

4.2.2. Ground Truth Liar

For the evaluation of the generated DSMs in the Netherlans, The Actueel Hoogtebestand Nederland (AHN) is used[1]. It is an ongoing collaborative initiative involving water boards, provinces, and Rijkswaterstaat, designed to create a comprehensive digital elevation model of the Netherlands. The program employs airborne laser scanning technology, which allows for the collection of detailed and precise height measurements across the entire country.

The most recent dataset, AHN4, was gathered between 2020 and 2022. It boasts a maximum standard deviation of 0.05m and a maximum systematic deviation of 0.05m, with a point density of approximately 10-14 points per square meter[1]. Research indicates that the reference dataset should be at least three

times more accurate than the DEM being evaluated. /citehawker2019accuracy. In this research, the estimated quality is expected to be less than three times as high, making this a viable option to use. For the purposes of this study, the AHN4 DEM was specifically cropped to fit the geographical boundaries of two study areas: The Hague and Delft.

Figure 4.5 shows the cropped area of the AHN4 DEM used for Delft and Den Haag. These images provide a visual reference to the specific regions of interest within the broader national dataset.

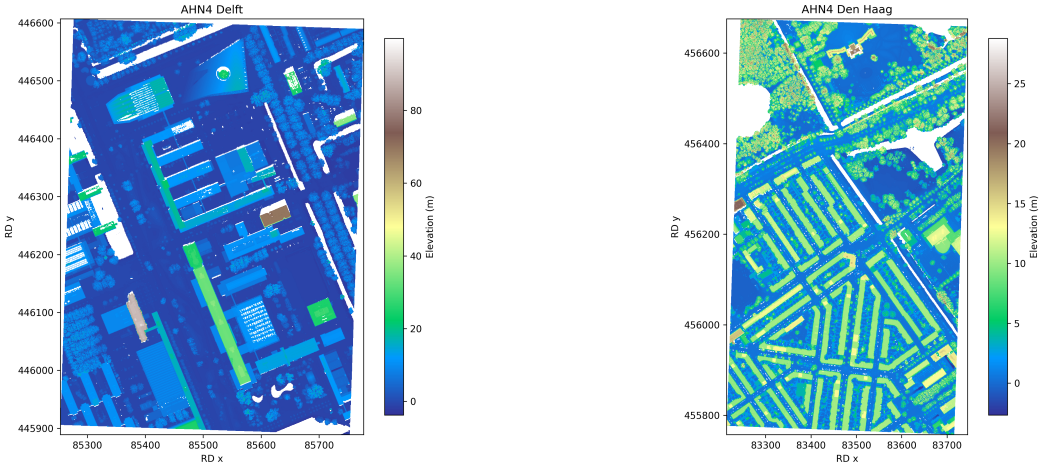


Figure 4.5: AHN4 DSM cropped for the area in Delft (left) and Den Haag (right).

5

Results

The results chapter systematically presents the outcomes of the experiments conducted using the advanced photogrammetric pipeline developed in this thesis. This chapter discusses the setup, execution, and results of the experiments designed to assess the efficacy of Digital Elevation Models (DEMs) derived from satellite imagery. Through meticulous adaptation of traditional photogrammetric methods and integration of innovative feature detection and matching algorithms, this research aimed to refine the process of DEM generation. The subsequent sections detail the software tools employed, the computational environment, and the specific input parameters for each dataset.

5.1. Experiment Setup and Input Parameters

For the reconstruction of Digital Elevation Models (DEMs) from satellite imagery, a comprehensive software stack was employed. The configuration details are as follows:

- **Python Version:** The pipeline was tested and operated using Python 3.8, ensuring compatibility and stability across all used libraries[32].
- **Image Processing Library:** The freeimage plugin for imageio was utilized to handle various image formats, enhancing the preprocessing capabilities within the Python environment[5].
- **DEM Reconstruction Software:** The experiment utilized two primary tools: ColmapForVisSat-Patched [19] and VisSatSatelliteStereo [20]. The software was complemented by VisSatSatelliteStereo for stereo image processing and DEM generation[43, 4].
- **Additional Libraries and Tools:**
  - **GDAL:** Geospatial Data Abstraction Library (GDAL) version 3.3.2 was used for its robust handling of geospatial data[11].
  - **Meshlab:** Meshlab-2020.09 was used for final mesh processing and analysis, providing essential tools for mesh editing and quality control[42].

All software installations and configurations were managed within a conda environment, which was configured to ensure all dependencies were correctly aligned and environmental variables set for optimal software interaction[2].

5.1.1. Computational Environment

The computational experiments were conducted on a dedicated device to ensure the processing power was sufficient for the demands of satellite image analysis and DEM reconstruction. The details of the computational setup are as follows:

- **Processor:** Intel(R) Core(TM) i7-10850H CPU @ 2.70GHz, 2.71 GHz
- **Installed RAM:** 32.0 GB (31.6 GB usable)
- **System Type:** 64-bit operating system, x64-based processor

- **Operating System:** Conducted within a Windows Subsystem for Linux (WSL) environment, leveraging its capabilities for UNIX-like scripting and toolchain integration.

This hardware configuration provided the robust performance required for the intensive computations involved in 3D modeling from high-resolution satellite images. The use of WSL allowed for the integration of Linux-based tools and libraries seamlessly within the Windows operating environment.

5.1.2. Input parameters

The datasets utilised in this thesis cover a variety of geographic locations, each with unique characteristics essential for photogrammetric analysis. This section provides a summarised overview of the input settings for each dataset, complementing the detailed descriptions provided in the Data Description chapter.

San Fernando, Argentina

The datasets from Argentina are part of the IARPA MVS3DM challenge[39], designed to enhance research in multi-view stereo photogrammetry. The datasets consist of three sites, each selected for its unique terrain features and data quality. The specifics of each site are presented in the table below:

Table 5.1: Input Settings for the Argentina Datasets

Site	Zone info	Easting (m)	Northing (m)	Width $\times$ Height (m)	Altitude Range (m)
Site 1	21S	354035	6182717	745 $\times$ 682	-20 to 100
Site 2	21S	354984	6185506	366 $\times$ 350	-20 to 80
Site 3	21S	355028	6187389	426 $\times$ 397	-20 to 90

Each site's Easting and Northing coordinates are provided in Universal Transverse Mercator (UTM) units. A useful tool for converting latitude and longitude to UTM is available at `https://www.latlong.net/lat-long-utm.html`.

Delft and The Hague, Netherlands

The Dutch datasets provide an urban contrast to the Argentine environments. They include detailed imagery from Delft and The Hague. The input settings for these locations are summarised in the table below:

Table 5.2: Input Settings for the Netherlands Datasets

Location	Zone info	Easting (m)	Northing (m)	Width $\times$ Height (m)	Altitude Range (m)
Delft	31N	594164	5762309	500 $\times$ 700	20 to 140
The Hague	31N	591804	5772304	500 $\times$ 900	20 to 150

These datasets were chosen for their high-resolution imagery and the availability of comprehensive ground truth data through the AHN4 LiDAR dataset[1], making them suitable for advanced photogrammetric analysis and DSM validation.

This section aims to provide a clear and concise overview of the dataset specifics used in the experiments, linking back to the more detailed descriptions available in the data chapter. This ensures that the experimental setup and results discussed in subsequent sections are grounded in a thorough understanding of the data characteristics and input settings.

5.1.3. Results from San Fernando, Argentina

This section presents the results obtained from applying both the original and adapted photogrammetric pipelines to the benchmark dataset from San Fernando, Argentina. The results showcase the baseline capabilities and improvements achieved through the adapted pipeline, utilizing high-resolution satellite imagery to produce detailed Digital Elevation Models (DEMs).

Visual Representation

The DEMs produced for each of the three sites are displayed below. Each row represents a different site and includes a side-by-side comparison with the input image, the DEM from the original pipeline, the DEM following adaptations with LightGlue and DISK features, and a color bar for height reference.

The first column of the figures (Figure 5.1, Figure 5.2, Figure 5.3) presents examples of the input images to provide an idea of the surface. The second column shows the outcome DSM of the original pipeline, while the third column displays the outcome of the adapted pipeline with LightGlue and DISK features.

Overall, the reconstruction is very good, despite some room for improvement at building boundaries and weakly-textured regions (e.g., ground and building tops) for both DSM and all sites. The resulting DEMs successfully capture key topographical and man-made features across all sites. Buildings, streets, and vegetation are clearly visible, contributing to a richly textured representation of the urban landscape. Even finer details such as light poles are discernible in Site 2, which underscores the precision of the photogrammetric methods employed using WorldView-3 imagery.

The DEMs from SV1 data show several gaps and uneven surfaces, especially over buildings. Some gaps match the locations of water bodies, which are difficult to handle with photogrammetry. However, there are more widespread gaps throughout the SV1 DEMs compared to those from WV3. These issues arise partly because of how the DEMs are created: missing data points are filled by estimating values from nearby data, but this method doesn't always perform well in areas with sparse data, like the spaces between tall buildings or open areas with little surface texture.

Moreover, the quality differences between SV1 and WV3 DEMs are also due to the satellite images themselves. For the SV1 datasets, we used only 21 images per site with a resolution of 0.5 meters, whereas the WV3 datasets used 50 images with a finer resolution of 0.3 meters. Having fewer images with lower resolution makes it harder to get detailed and accurate DEMs.

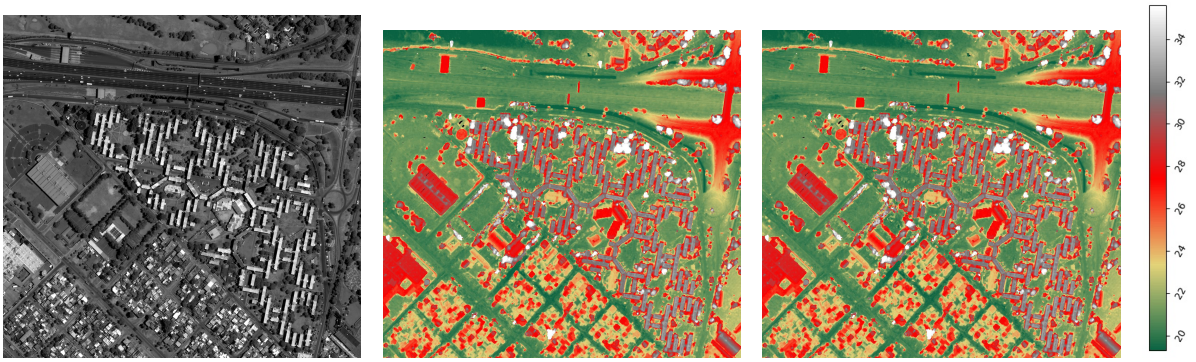


Figure 5.1: From left to right: Input image, initial DEM, adapted DEM, and height [m] color bar for Site 1.

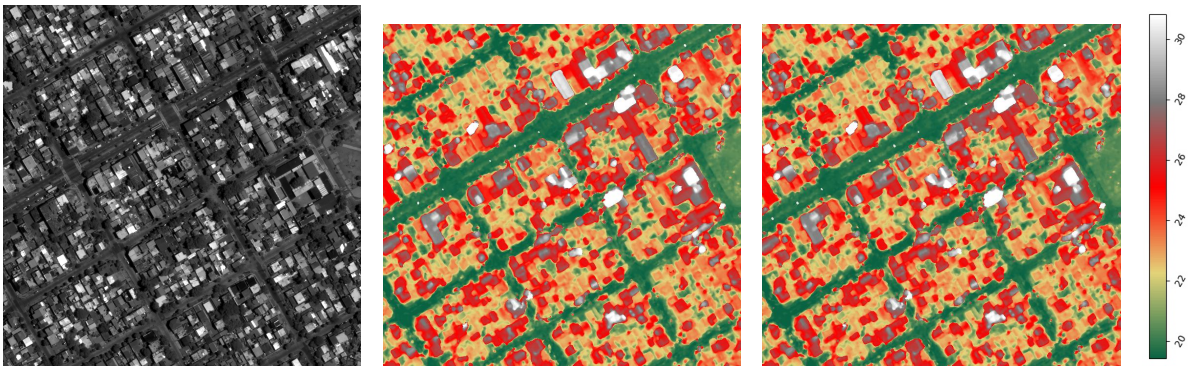


Figure 5.2: From left to right: Input image, initial DEM, adapted DEM, and height [m] color bar for Site 2.

Figure 5.3: From left to right: Input image, initial DEM, adapted DEM, and height [m] color bar for Site 3.

5.1.4. Results from the Netherlands

This section presents the results obtained from applying both the original and adapted photogrammetric pipelines to the datasets from Delft and The Hague, Netherlands. These results illustrate the improvements achieved by adapting the pipeline to handle high-resolution SuperView-1 satellite imagery effectively.

Delft

Initial DEM Results The initial Digital Surface Model (DSM) produced for Delft using the original pipeline settings is shown below, followed by the results from the adapted pipeline.

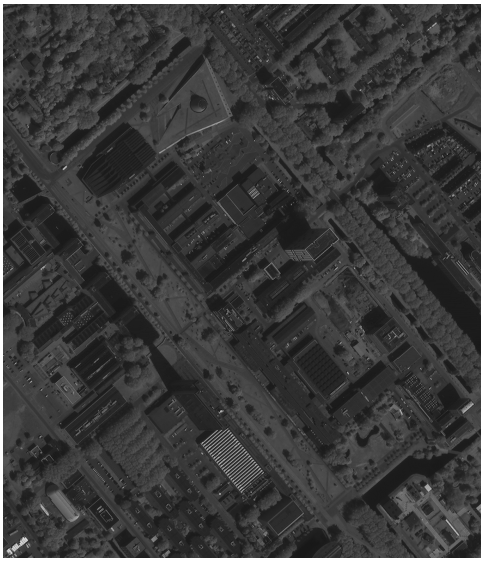
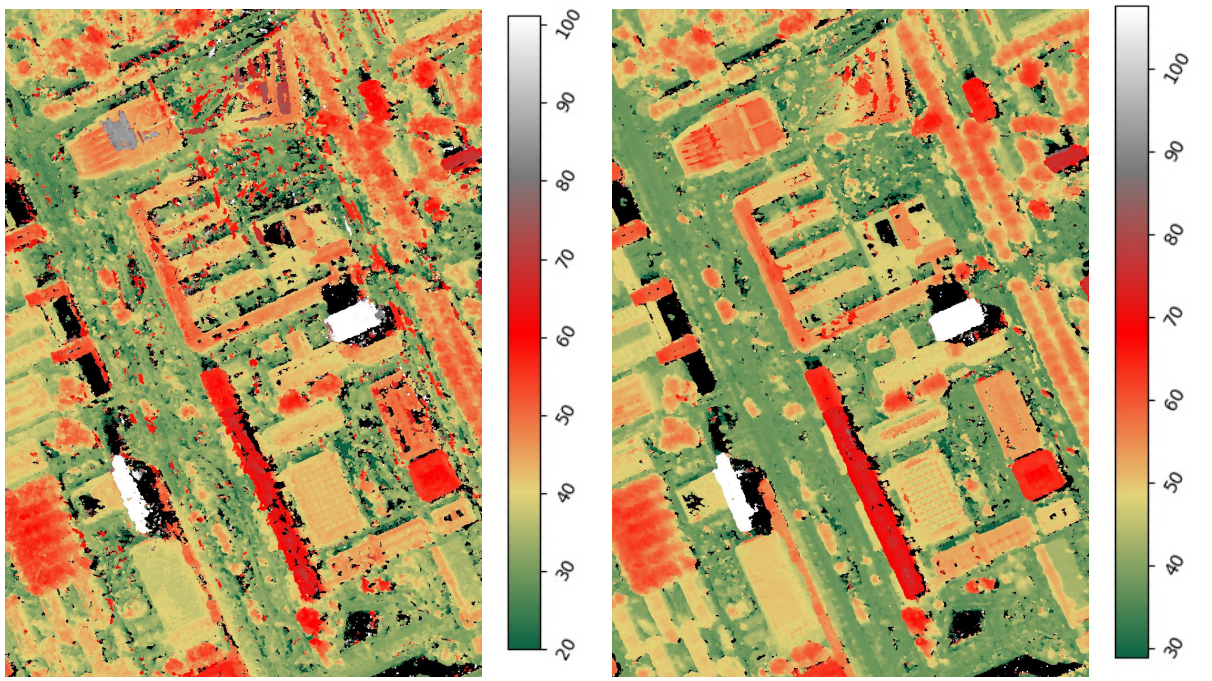


Figure 5.4: Input satellite image of Delft.



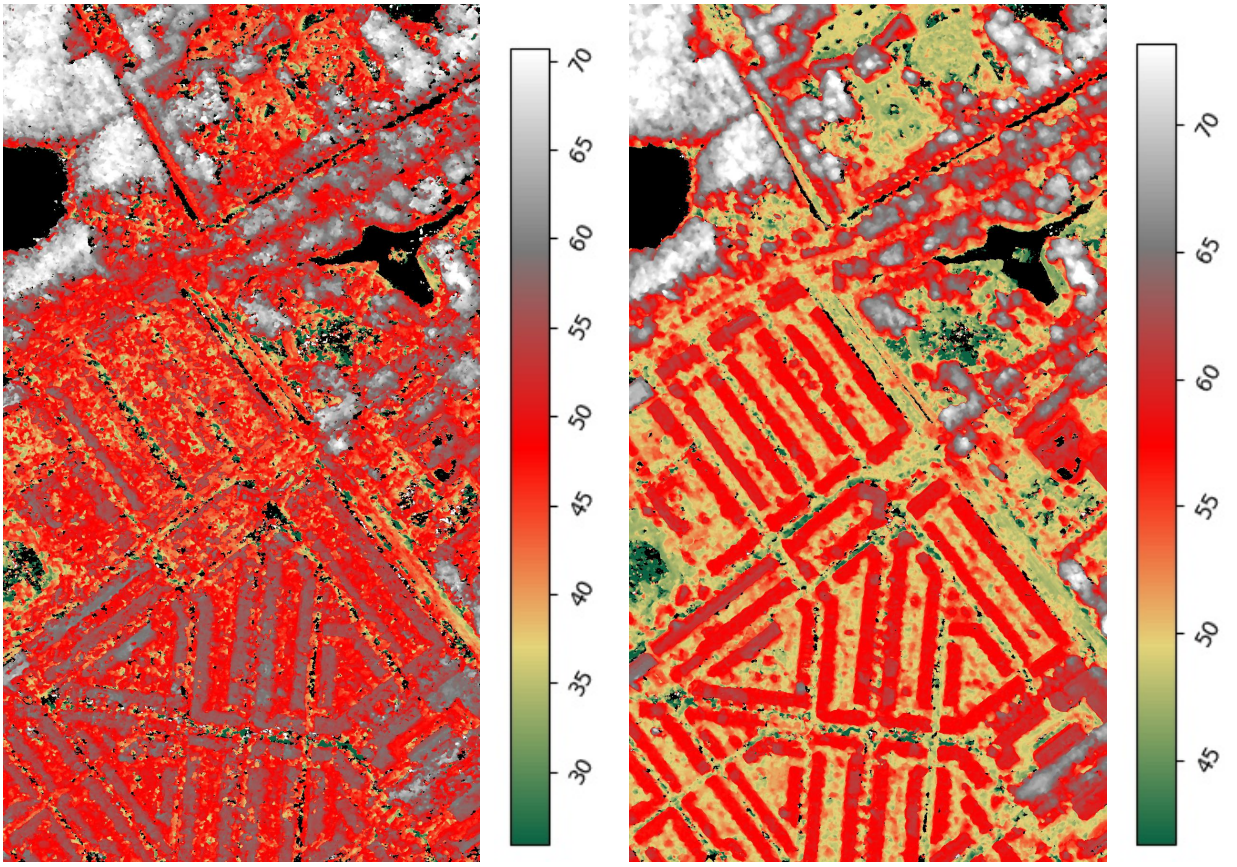
(a) DSM from original pipeline **(b)** Height (m) **(c)** DSM from adapted pipeline **(d)** Height (m)

Figure 5.5: Comparison of the original and adapted DSMs for Delft.

Den Haag

Initial DEM Results Similarly, the initial DEM results for Den Haag are presented, showcasing the effectiveness of the original pipeline followed by the enhancements made in the adapted version.

Figure 5.6: Example of an input satellite image of Den Haag.



(a) DSM from original pipeline **(b)** Height (m) **(c)** DSM from adapted pipeline **(d)** Height (m)

Figure 5.7: Comparison of the original and adapted DSMs for Den Haag.

Adaptation for SuperView-1 Datasets

After modifying the photogrammetric pipeline to process SuperView-1 (SV1) satellite images, the resulting Digital Elevation Models (DEMs) can be seen in Figures 5.5 and 5.7. When you look at these next to the DEMs from the WorldView-3 (WV3) images in Figures 5.1, 5.2, and 5.3, you can see that the models are not at all the same in terms of how smooth and complete they are.

The DEMs from SV1 data show several gaps and uneven surfaces, especially over buildings. Some gaps match the locations of water bodies, which

5.2. Evaluation of Digital Surface Models (DSMs)

This section evaluates the accuracy and quality of the Digital Surface Models (DSMs) produced by the original and adapted photogrammetric pipelines, focusing solely on the dataset from the Netherlands due to the minimal differences observed in the WorldView-3 (WV3) dataset results and time constraints.

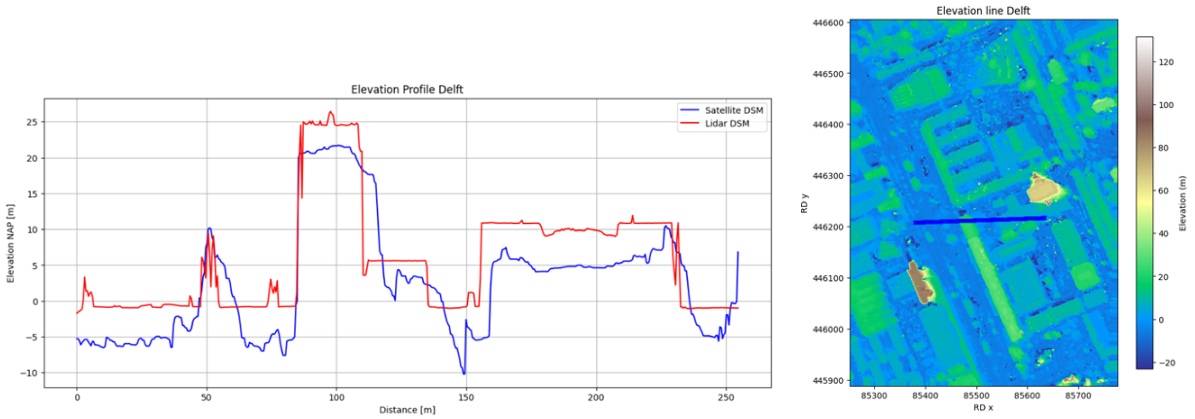
5.2.1. Vertical Co-registration

The DSMs generated from the satellite pipeline still exhibit some gaps and open areas. These gaps are filled using the nearest neighbor method [**method**]. Filling these gaps is crucial to ensure a continuous surface model, which is necessary for various geospatial analyses and applications. The alignment accuracy is assessed by comparing the DSMs against high-resolution LiDAR data, serving as the ground truth.

To better understand the vertical alignment of the DSMs, two transects were created for Delft and Den Haag. These transects were designed to cover the most representative features in the images, including buildings, roads, and open spaces.

For Delft, the DSM created using the adapted pipeline with the transect is shown in Figure 5.8b. The

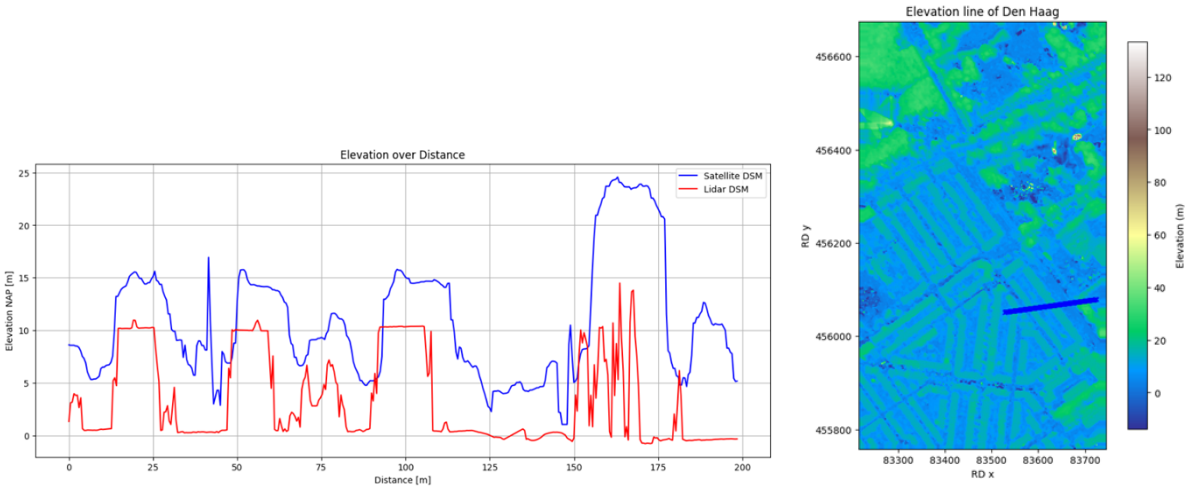
transect itself is show in figure 5.8a and clearly shows a vertical offset where the created DSM lies lower than the LiDAR ground truth.



(a) (b)

Figure 5.8: Left: Vertical profile across the intersection. Right: Intersection visualized on the interpolated DSM.

The DSM created using LightGlue with the ransect for Den Haag is shown in Figure 5.9b. The vertical profile of the transect is plotted in figure 5.9a and this transect indicates a vertical offset where the created DSM is higher than the LiDAR ground truth.

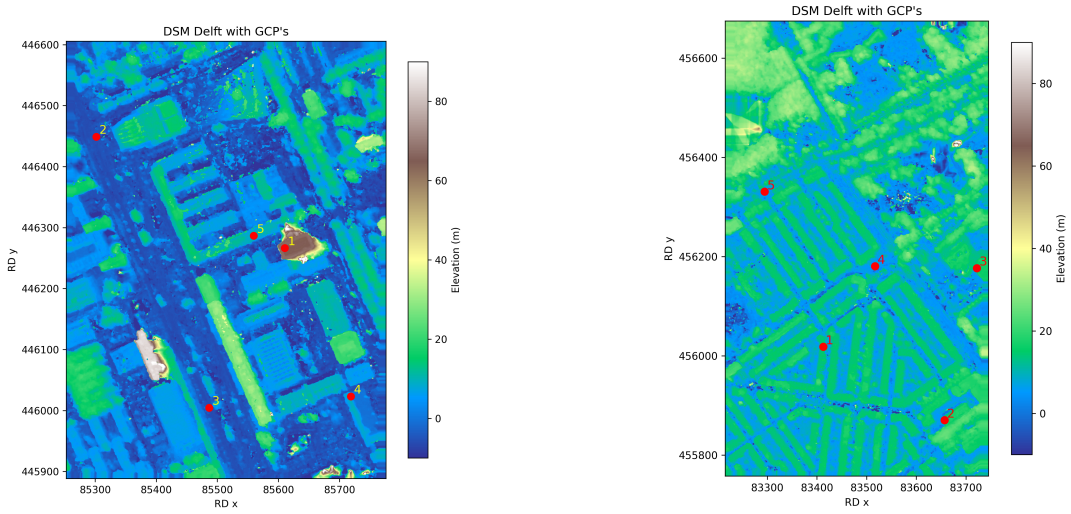


(a) (b)

Figure 5.9: Left: Vertical profile across the intersection. Right: Intersection visualized on the interpolated DSM for Den Haag.

These transects clearly illustrate the presence of vertical offsets in the results. In Delft, the created DSM is positioned lower than the LiDAR ground truth, whereas in Den Haag, it is positioned higher.

Visually there can also be seen a horizontal bias but this is small and outside the scope of this study, this thesis focuses largely on the vertical offset in the DSMs. However, future work could extend this to include horizontal alignment corrections. The vertical co-registration process follows the methodology described in the method subsection 3.6.3 . This process involves several steps: Five points were selected in the LiDAR dataset that appeared to be best reconstructed in the DSM. These points are primarily located on buildings and roads to simplify the isolation of vertical offsets, given that horizontal alignment is not performed. For the DSM created using the adapted pipeline for Delft and Den Haag, is illustrated in figure 5.10, showing the locations used for the vertical co-registration in Delft.



(a) Delft

(b) Den Haag

Figure 5.10: DSM's of the adapted pipeline for Delft and Den Haag with marked points used for the vertical allignment.

The vertical difference between the generated DSM and the ground-truth LiDAR DSM was calculated at each selected point. This is done for the DSM generated from both the original pipeline and the adapted one. The coordinates along the vertical differences between the original DSM, the adapted DSM, and the LiDAR ground truth are presented in Table 5.3 for Delft and Table 5.4 for Den Haag.

Table 5.3: Vertical Differences at Selected Points for Delft

Point	Location (RD)	Original DSM (m)	Adapted DSM (m)	Ground Truth DSM (m)
1	(85610.50, 446266.30)	58.68	64.99	69.61
2	(85301.90, 446448.90)	-15.49	-4.88	-0.96
3	(85486.50, 446004.50)	-12.02	-6.16	-1.03
4	(85718.50, 446023.00)	1.27	6.64	4.62
5	(85559.60, 446286.60)	3.37	9.82	15.01

Table 5.4: Vertical Differences at Selected Points for Den Haag

Point	Location (RD)	Original DSM (m)	Adapted DSM (m)	Ground Truth DSM (m)
1	(83412.69, 456018.36)	14.06	14.79	10.19
2	(83656.84, 455870.82)	15.10	18.31	12.90
3	(83721.95, 456176.92)	10.03	15.94	11.91
4	(83516.91, 456180.66)	-1.55	4.91	0.33
5	(83294.68, 456331.08)	2.91	6.62	0.92

The vertical differences highlighted in Tables show significant discrepancies between the DSM heights obtained from the original and adapted pipelines compared to the ground-truth LiDAR data. All heights are referenced against NAP, and the coordinates are in the RD coordinate system, ensuring that the data align with local geospatial standards.

Mean Offsets and Adjustments

The mean vertical offsets calculated for Delft and Den Haag illustrate the systematic bias present in the DSMs produced by both the original and adapted pipelines. For Delft, the mean difference from the original DSM to ground truth is approximately 10.29 meters, while for the adapted DSM, it reduces

to about 4.17 meters. Similarly, for Den Haag, the original DSM shows a mean difference of 2.36 meters, which increases slightly to 4.86 meters with the adapted DSM, suggesting that while the adapted pipeline improved vertical accuracy in Delft, it needs further refinement in Den Haag.

These mean differences are significant as they guide the vertical co-registration process, which aims to eliminate these offsets to align the DSMs closely with the ground truth. This alignment process is described in the method subsection 3.6.3.

5.2.2. Numerical Evaluation

Following the application of vertical adjustments, the next step involves a comprehensive numerical evaluation of the adjusted DSMs. This evaluation will quantify the improvements in terms of standard error metrics such as RMSE (Root Mean Square Error), Mean Absolute Error, and others, as outlined in the method section 3.6.4. This stage is crucial for validating the effectiveness of the vertical co-registration process and for providing a quantifiable measure of the enhancements achieved through the adaptations in the photogrammetry pipeline.

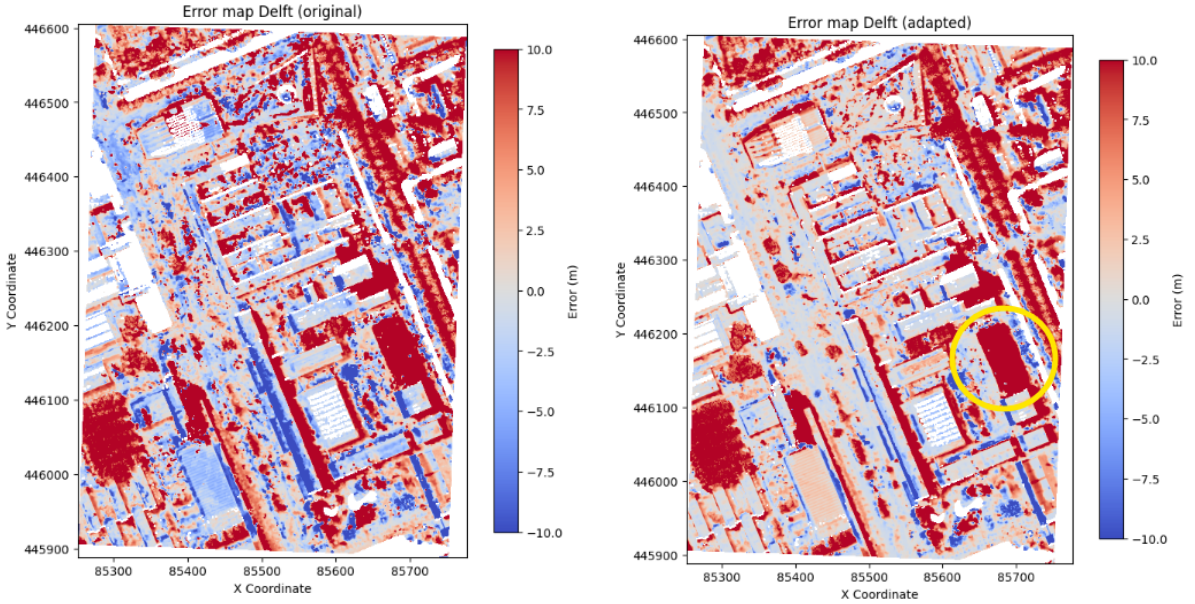
5.3. Numerical Evaluation of Digital Surface Models

This section presents a detailed numerical analysis of the DSMs produced using both the original and adapted photogrammetry pipelines for Delft and Den Haag. The evaluation focuses on key statistical metrics such as Mean Difference, Standard Deviation, and Root Mean Square Error (RMSE), which quantify the accuracy and reliability of the DSMs in comparison to high-resolution LiDAR ground truth data.

The error analysis involves comparing the DSM elevation values directly against the corresponding LiDAR ground truth. The error calculated (DSM value - Ground truth) highlights the discrepancies and shows where the DSM either overestimates or underestimates the actual terrain features. Notably, areas without LiDAR data are represented in white on the error maps, indicating regions where the evaluation could not be conducted.

Error Visualization Delft

Visual comparisons for Delft are shown in figure 5.11, illustrating the error distribution for both the original and adapted DSMs.

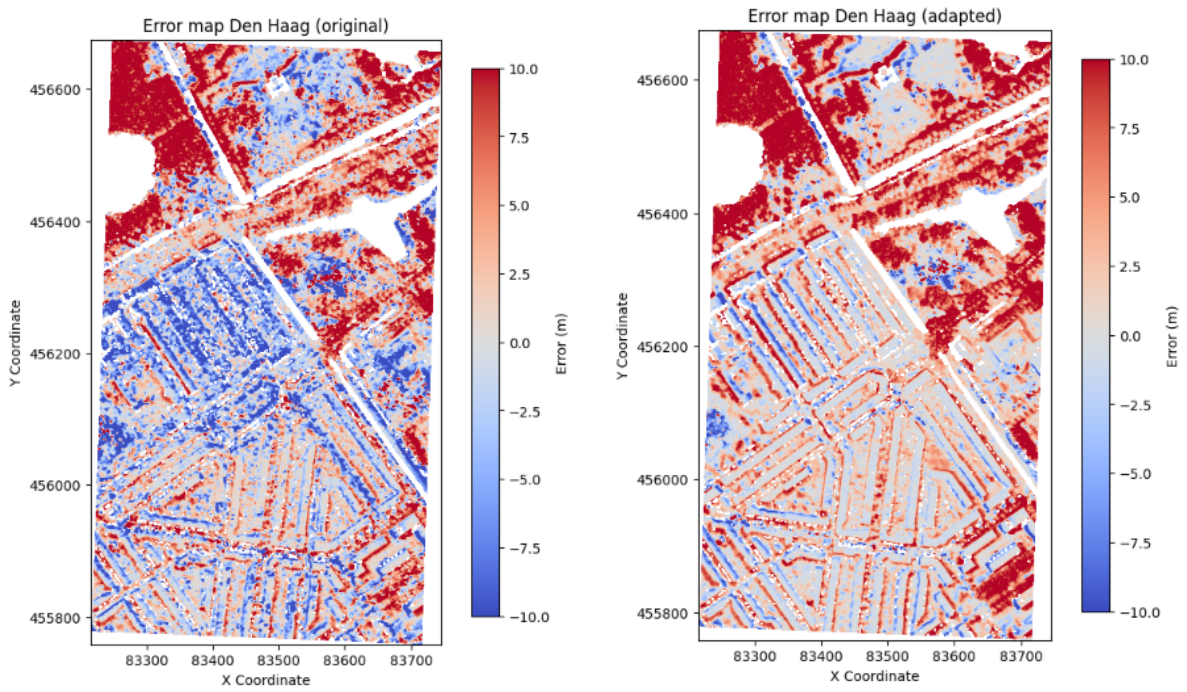


(a) Original Pipeline **(b)** Adapted Pipeline

Figure 5.11: Error maps for DSMs of Delft comparing to LiDAR ground truth.

Error Visualization Den Haag

Similar error distributions are shown for Den Haag in figure 5.12, comparing the original and adapted DSMs.



(a) Original Pipeline **(b)** Adapted Pipeline

Figure 5.12: Error maps for DSMs of Den Haag comparing to LiDAR ground truth.

5.3.1. Statistical Error Analysis

The statistical evaluation includes detailed metrics for each city and pipeline configuration. The tables below summarize the calculated Mean Difference, Standard Deviation, and RMSE, reflecting the accuracy and precision of the DSMs.

Statistical Metrics for Delft

Table 5.5: Statistical Metrics for Delft DSMs

Pipeline	Mean Difference (m)	Standard Deviation (m)	RMSE (m)
Original	3.72	13.09	13.60
Adapted	3.52	9.60	10.22

Statistical Metrics for Den Haag

Table 5.6: Statistical Metrics for Den Haag DSMs

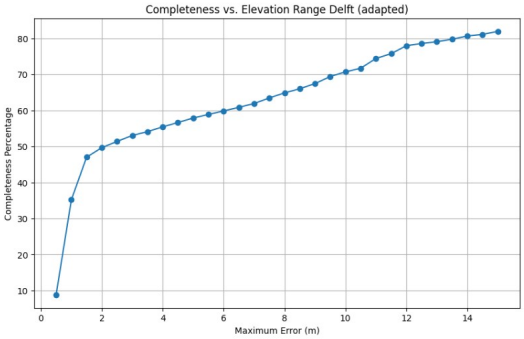
Pipeline	Mean Difference (m)	Standard Deviation (m)	RMSE (m)
Original	0.88	8.21	8.25
Adapted	3.25	6.27	7.06

Note: Discuss the implications of these results, focusing on how the adapted pipeline improves the DSM accuracy and what could be causing the remaining discrepancies, such as the large errors near trees and high buildings in the adapted Delft model.

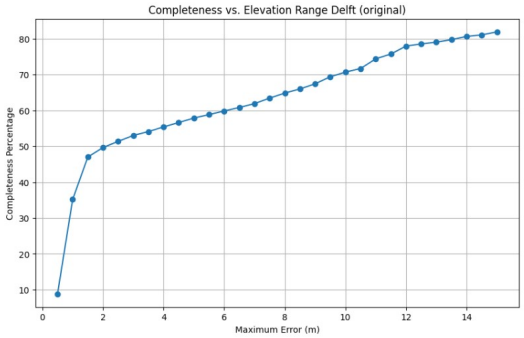
Completeness Analysis

The completeness analysis is a crucial component of the evaluation, as it assesses how thoroughly the Digital Surface Models (DSMs) represent the terrain under various error thresholds. This analysis helps in understanding the proportion of the terrain that is accurately captured within specific error margins, providing insights into the effectiveness of the DSM reconstruction in capturing terrain features.

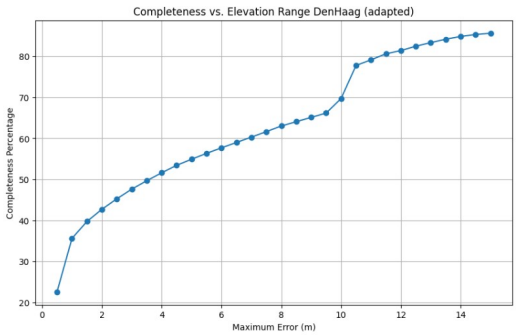
Completeness is quantified by the percentage of the DSM that falls within predefined error thresholds relative to the LiDAR ground truth. This measure is indicative of how much of the terrain is modeled within acceptable error limits, which is essential for applications relying on the precision of elevation data. The analysis is visualized through four plots in figure 5.13, each corresponding to one of the pipeline configurations for Delft and Den Haag.



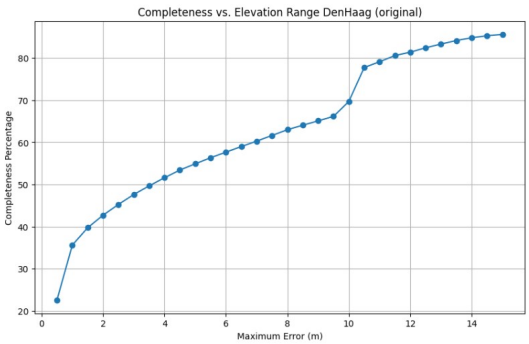
(a) Completeness of Adapted DSM for Delft



(b) Completeness of Original DSM for Delft



(c) Completeness of Adapted DSM for Den Haag



(d) Completeness of Original DSM for Den Haag

Figure 5.13: Plots of the Completeness distribution of Delft And Den Haag

5.4. Impact of Replacing SIFT with DISK and LightGlue Matching

This section delves into the significant modifications made to the photogrammetric pipeline, specifically the adoption of DISK features and LightGlue matching as replacements for the traditional SIFT feature detection and matching. The analysis leverages the Argentinian dataset, which provides a diverse terrain and feature set, ideal for evaluating the efficacy of these advanced matching techniques.

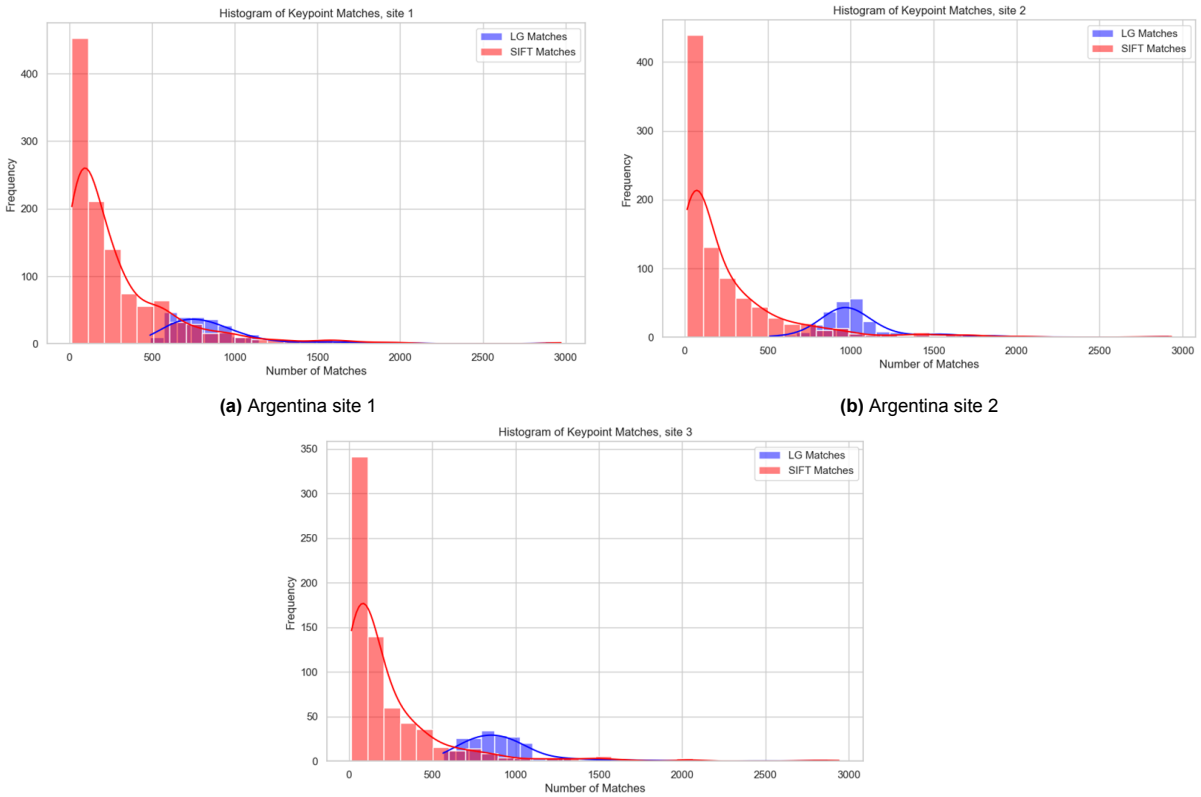
5.4.1. Dataset and Methodology

The Argentinian dataset was chosen for its varied landscape, which challenges the feature matching capabilities of photogrammetric algorithms. Using this dataset allows for a detailed comparison of how DISK and LightGlue enhance feature detection and matching compared to the traditional SIFT approach. Both SIFT and DISK methods were configured to detect 3000 initial features per image, ensuring consistency in the comparative analysis and isolating improvements to the efficacy of feature matching itself.

Feature Matching Performance

Initial results, as illustrated in Figure 5.14, show the distribution of matched features across different image pairs. It is evident that DISK with LightGlue matching yields fewer matched image pairs, indi-

cating a more selective matching process. This selectivity stems from a preemptive matching step where only the eight most promising matches for each image are pursued. This strategy enhances the pipeline's efficiency and robustness by excluding less reliable matches that could degrade the quality of the sparse reconstruction.



(a) Argentina site 1

(b) Argentina site 2

(c) Argentina site 3

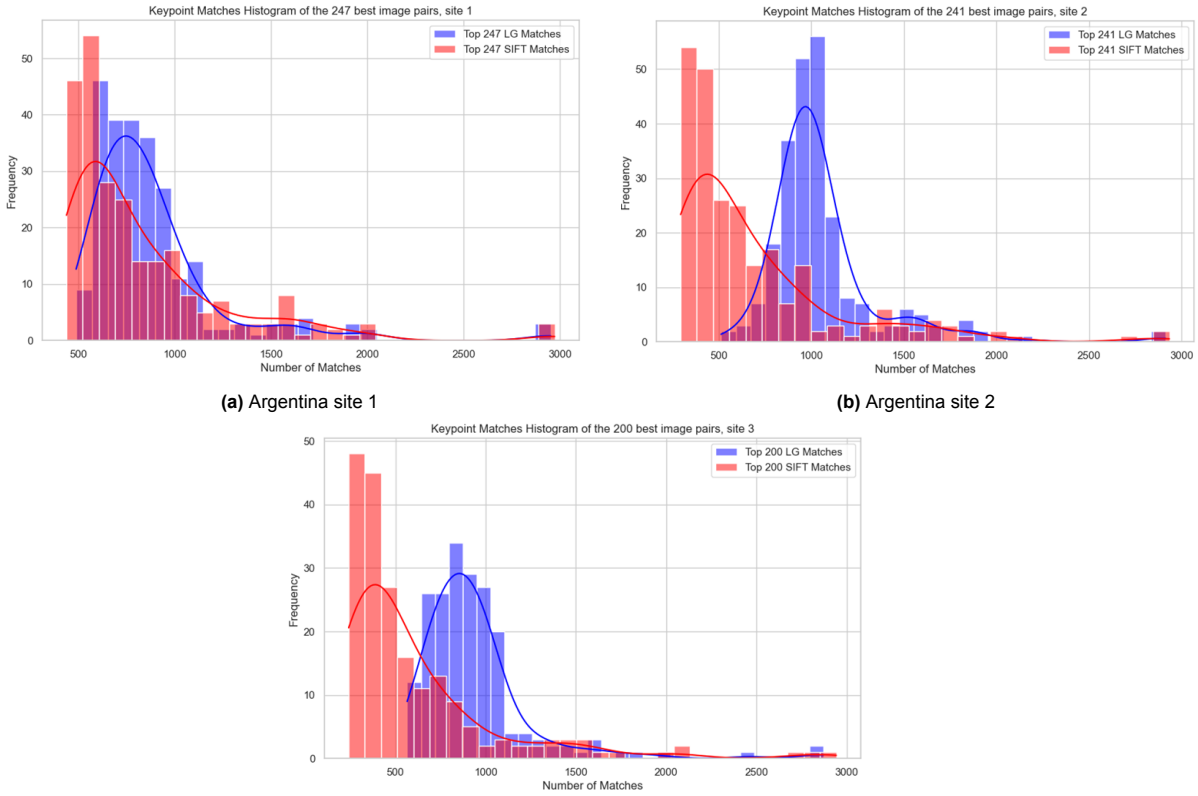
Figure 5.14: Histogram of the matched keypoints for the Argentina datasets for Sift (Red) and Lightglue (Blue)

SIFT Matching Shortcomings

SIFT, in contrast, demonstrates a broader spread of matches across image pairs, including many pairs with very few matches. These low-match pairs, while numerous, often contribute less to the quality of the 3D reconstruction and can potentially introduce noise into the dataset.

Detailed Comparative Analysis

For a more focused comparison, taking into account only the top most used image pairs, Figure 5.15 zooms into the higher-quality matches, setting the threshold at the level of matches typically achieved by LightGlue. Here, it becomes apparent that the matches retained by LightGlue not only are fewer in total image pairs but also are of higher quality in terms of the number of keypoints matched per image pair. This indicates a more selective yet effective matching strategy, which is likely to contribute positively to the subsequent stages of the photogrammetric process.



(a) Argentina site 1

(b) Argentina site 2

(c) Argentina site 3

Figure 5.15: Histogram of the best matched keypoints for the Argentina datasets for Sift (Red) and Lightglue (Blue)

Implications for Sparse Reconstruction

The higher quality of matches obtained with DISK and LightGlue suggests that these matches are more likely to be correct, which is crucial for generating a reliable and accurate sparse point cloud. The precision in matching directly impacts the quality of the 3D models produced, potentially leading to better surface detail and higher fidelity in the reconstructed models.

6 Discussion

This chapter revisits the research questions introduced in the introduction and discusses the findings from the experimental chapters. It evaluates the relevance, effectiveness, and limitations of Digital Terrain Models (DTMs) derived from satellite imagery.

6.1. Addressing the Research Questions

General Research Question: How can satellite imagery be used to derive high-quality and reliable Digital Terrain Models for diverse applications?

6.1.1. Relevance and Benefits of DTMs

Satellite-derived DTMs are invaluable for areas difficult to access, offering a globally available and cost-effective alternative to traditional survey methods. The increasing quality and availability of satellite imagery make it a crucial tool in remote sensing and environmental monitoring.

6.1.2. Available Data and Requirements

The study utilized datasets from SuperView-1 and WorldView-3 satellites, which include an RPC model and multiple image sets for specific locations. The successful adaptation of these datasets demonstrates that the photogrammetric pipeline can be expanded to a broader range of satellite data, provided sufficient RPC information and image overlap exist.

6.1.3. Efficient Workflow for DTM Generation

The transition from traditional photogrammetric techniques to those optimized for satellite imagery has significantly enhanced the efficiency and accuracy of DTM production. The workflow includes the following steps:

- Use of RPC models to account for the complex camera models used in satellite imagery.
- Implementation of image preprocessing steps such as skew correction and cropping to focus on the area of interest, thereby reducing computational load.
- Adaptations from traditional pipelines to handle the unique characteristics of satellite data, such as adjusting for varying lighting conditions and geometric distortions.

6.1.4. Interaction of GCPs with Satellite Imagery

GCPs are used by sampling from the high-resolution LiDAR dataset, which is necessary to filter out large offsets and enhance the DTM quality. This integration is crucial for aligning the satellite-derived models closely with the ground truth, thereby improving the accuracy of the final product.

6.1.5. Validation Against Benchmarks or Ground Truths
DTMs were aligned and compared against high-resolution LiDAR data, which revealed consistent offsets influenced by satellite image quality and RPC model precision. This rigorous validation process underscores the importance of accurate calibration in photogrammetric workflows.

6.1.6. Susceptibility to Quality Issues
Performance variations among datasets necessitated specific optimizations to address unique challenges. Replacing SIFT with DISK and LightGlue significantly mitigated deficiencies related to image distortions and environmental factors affecting image quality.

6.2. Limitations and Implications
6.2.1. Methodological Limitations
Despite advancements, the photogrammetric pipeline demands high-quality images and significant computing power, limiting its application in resource-constrained settings. The absence of GCP integration in this study points to a significant gap in achieving the highest geospatial accuracy.

6.2.2. Practical Implications
The enhanced DTMs offer significant promise for applications such as urban planning and environmental management. However, the difficulty in feature detection in open areas remains a challenge even with advanced algorithms. This limitation underscores the need for ongoing improvements in algorithmic approaches to feature detection.

6.2.3. Comparison and Contextualization with Other Studies
- The overall vertical accuracy of our method is comparable to other DEMs constructed using different technologies, with certain biases and errors characteristic of the imaging and processing methods used.
- The presence of systematic biases in height estimation and the challenges posed by different terrain types (such as slopes and flat areas) align with findings from other studies, indicating common challenges across different DEM construction methodologies.

6.3. Further Research
Further research could explore the integration of more robust GCP usage, enhanced algorithms for feature detection in open areas, and more refined methods for handling varying terrain types. Additionally, exploring other interpolation methods could potentially improve the accuracy and quality of the generated DTMs.

This discussion synthesizes the insights gained and sets the stage for future advancements in the field of satellite-derived digital terrain modeling.

7

Conclusion and Recommendations

This thesis has explored the development and validation of Digital Terrain Models (DTMs) derived from high-resolution satellite imagery, demonstrating the potential and limitations of advanced photogrammetric techniques in capturing detailed terrestrial features.

7.1. Conclusion

The research conducted has provided significant insights into the use of satellite imagery for DTM production. Through the adaptation of advanced photogrammetric algorithms, notably DISK for feature detection and LightGlue for feature matching, the study has shown that it is possible to enhance the accuracy and reliability of DTMs. These models are crucial for environmental monitoring, urban planning, and other geospatial applications, especially in areas where traditional survey methods are impractical.

Key findings of this thesis include:

- The adapted photogrammetric pipeline significantly improved the vertical accuracy of DTMs compared to traditional methods, as evidenced by comparisons with high-resolution LiDAR data.
- Despite these improvements, challenges remain in areas with sparse features, such as open landscapes or water bodies, where feature detection algorithms still struggle to find sufficient matches.
- The study highlighted the importance of Ground Control Points (GCPs) sampled from LiDAR data, which were crucial in mitigating large vertical offsets and aligning the satellite-derived models with ground truth.
- The vertical co-registration process proved essential for correcting systematic biases and discrepancies in elevation data derived from satellite imagery.

Ultimately, this thesis underscores the transformative potential of satellite-derived DTMs, while also acknowledging the critical need for further enhancements in photogrammetric techniques to address the identified limitations.

7.2. Future Work

The findings from this thesis pave the way for several avenues of future research, which could further refine the utility and accuracy of DTMs derived from satellite imagery:

- **Integration of Machine Learning:** Future studies could explore the use of more sophisticated machine learning algorithms to improve feature detection and matching, especially in challenging environments like urban sprawls or heavily vegetated areas.
- **Enhanced GCP Integration:** Investigating automated methods for GCP selection and integration could help in reducing the manual effort required and improving the geospatial accuracy of the resulting DTMs.

- **Multi-Sensor Data Fusion:** Combining data from multiple satellite sensors or integrating satellite data with aerial or drone-based sensors might provide richer datasets for DTM generation, offering higher resolution and better error mitigation.
- **Algorithmic Adjustments for Open Areas:** Developing specialized algorithms that can effectively handle feature-poor regions could significantly enhance the completeness and accuracy of satellite-derived DTMs.
- **Longitudinal Studies:** Conducting studies over longer periods could help in understanding the temporal dynamics captured by DTMs, which is crucial for monitoring environmental changes and planning urban developments.

Continued research in these areas could lead to significant advancements in the field of remote sensing and digital elevation modeling, contributing to more reliable and comprehensive geospatial data resources.

The implications of this research extend beyond academic inquiry, influencing practical applications in civil engineering, urban planning, disaster management, and environmental monitoring. As satellite imaging technologies and photogrammetric techniques continue to evolve, the potential for their application in creating detailed, accurate, and accessible DTMs is bound to expand, offering profound benefits across multiple sectors.

References

[1] Actueel Hoogtebestand Nederland. *Over het programma ahn*. [Accessed: 2024]. 2024. URL: https://www.ahn.nl/over-het-programma-ahn.

[2] Inc. Anaconda. *Anaconda Software Distribution*. 2024. URL: https://anaconda.com.

[3] Marc Bosch et al. "A multiple view stereo benchmark for satellite imagery". In: *2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE. 2016, pp. 1–9.

[4] Sebastian Bullinger, Christoph Bodensteiner, and Michael Arens. "3D Surface Reconstruction from Multi-Date Satellite Images". In: *arXiv preprint arXiv:2102.02502* (2021).

[5] Imageio Contributors. *imageio: A Python library for reading and writing image data*. Version 2.27.0. 2024. URL: https://imageio.github.io/.

[6] Samuel Picton Drake. "Converting GPS coordinates [phi, lambda, h] to navigation coordinates (ENU)". In: (2002).

[7] European Space Agency. *WorldView-3 instruments*. Retrieved April 9, 2024. European Space Agency. n.d. URL: https://earth.esa.int/eogateway/missions/worldview-3#instruments-section (visited on 04/09/2024).

[8] Gabriele Facciolo, Carlo De Franchis, and Enric Meinhardt-Llopis. "Automatic 3D reconstruction from multi-date satellite images". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017, pp. 57–66.

[9] Wolfgang Förstner and Bernhard P Wrobel. *Photogrammetric computer vision, volume 11 of Geometry and Computing*. 2016.

[10] Simon Fuhrmann, Fabian Langguth, and Michael Goesele. "Mve-a multi-view reconstruction environment." In: *GCH* 3 (2014), p. 4.

[11] GDAL Development Team. *GDAL - Geospatial Data Abstraction Library*. Accessed: 2024-07-10. Open Source Geospatial Foundation. URL: https://gdal.org/.

[12] Zhang Guo and Yuan Xiuxiao. "On RPC model of satellite imagery". In: *Geo-spatial Information Science* 9.4 (2006), pp. 285–292.

[13] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[14] Heliguy. *Drones and DEMs vs DTMs vs DSMs*. [Photograph]. 2021. URL: https://www.heliguy.com/blogs/posts/drones-and-dems-vs-dtms-vs-dsms (visited on 04/10/2024).

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

[16] *INDIA'S JOURNEY TOWARDS EXCELLENCE IN BUILDING EARTH OBSERVATION CAMERAS*. Scientific Figure on ResearchGate. URL: https://www.researchgate.net/figure/Illustration-of-the-push-broom-scan-technique-t-is-the-integration-time-Reproduced_fig3_349052642 (visited on 06/10/2024).

[17] Michal Irani, P Anandan, and Daphna Weinshall. "From reference frames to reference planes: Multi-view parallax geometry and applications". In: *Computer Vision—ECCV'98: 5th European Conference on Computer Vision Freiburg, Germany, June 2–6, 1998 Proceedings, Volume II 5*. Springer. 1998, pp. 829–845.

[18] Michal Jancosek and Tomas Pajdla. "Multi-view reconstruction preserving weakly-supported surfaces". In: *CVPR 2011*. IEEE. 2011, pp. 3121–3128.

[19] Kai-46. *ColmapForVisSat*. GitHub repository. 2024. URL: https://github.com/Kai-46/ColmapForVisSat.

[20] Kai-46. *VisSatSatelliteStereo*. GitHub repository. 2024. URL: https://github.com/Kai-46/VisSatSatelliteStereo.

[21] Arno Knapitsch et al. “Tanks and temples: Benchmarking large-scale scene reconstruction”. In: *ACM Transactions on Graphics (ToG)* 36.4 (2017), pp. 1–13.

[22] Hugo Ledoux et al. *Computational Modelling of Terrains*. 2023.

[23] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. “Lightglue: Local feature matching at light speed”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 17627–17638.

[24] Rafał Mantiuk, Scott Daly, and Louis Kerofsky. “Display adaptive tone mapping”. In: *ACM SIGGRAPH 2008 papers*. 2008, pp. 1–10.

[25] MathWorks. *Transform geodetic coordinates to local east-north-up (ENU) coordinates*. Accessed: 2023-06-01. n.d. URL: https://www.mathworks.com/help/map/ref/geodetic2enu.html.

[26] ml31415. *numpy-groupies*. [Accessed: 10-Jun-2024]. 2024. URL: https://github.com/ml31415/numpy-groupies.

[27] Pierre Moulon, Pascal Monasse, Renaud Marlet, et al. “Openmvg. an open multiple view geometry library”. In: *An Open Multiple View Geometry Library* (2014).

[28] Christopher Nuth and Andreas Kääb. “Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change”. In: *The Cryosphere* 5.1 (2011), pp. 271–290.

[29] OpenCV. *OpenCV: Smoothing Images*. [Accessed: 10-Jun-2024]. 2024. URL: https://docs.opencv.org/3.4/d4/d13/tutorial_py_filtering.html.

[30] RJ Oudman. “The transformation of GPS into NAP heights”. In: (2006).

[31] pyproj. *pyproj - Projections and transformations using PROJ*. Accessed: 2023-06-01. n.d. URL: https://pyproj4.github.io/pyproj/stable/.

[32] Python Software Foundation. *Python*. Python Software Foundation. Beaverton, OR, 2019. URL: https://www.python.org/.

[33] ResearchGate. *Radial distortion (barrel and pincushion)*. Scientific Figure on ResearchGate. Accessed: 2024-06-09. June 2024.

[34] Paul-Edouard Sarlin et al. “Superglue: Learning feature matching with graph neural networks”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 4938–4947.

[35] Satellietdataportaal.nl. *Satellietdataportaal*. [Accessed: 10-Jun-2024]. 2024. URL: https://www.satellietdataportaal.nl.

[36] Johannes L Schonberger and Jan-Michael Frahm. “Structure-from-motion revisited”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 4104–4113.

[37] Johannes L. Schönberger and Jan-Michael Frahm. *COLMAP - General Camera Models*. [Accessed: 10-Jun-2024]. 2023. URL: https://colmap.github.io/database.html.

[38] Noah Snavely, Steven M Seitz, and Richard Szeliski. “Modeling the world from internet photo collections”. In: *International journal of computer vision* 80 (2008), pp. 189–210.

[39] SpaceNet. *IARPA Multi-View Stereo 3D Mapping Dataset*. [Accessed: 27-Jun-2024]. 2024. URL: https://spacenet.ai/iarpa-multi-view-stereo-3d-mapping/.

[40] Michał Tyszkiewicz, Pascal Fua, and Eduard Trulls. “DISK: Learning local features with policy gradient”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 14254–14265.

[41] Ostap Viniavskyi et al. “OpenGlue: Open source graph neural net based pipeline for image matching”. In: *arXiv preprint arXiv:2204.08870* (2022).

[42] Visual Computing Lab - ISTI - CNR. *MeshLab*. Version 2020.09 or older. Visual Computing Lab - ISTI - CNR. 2020. URL: https://www.meshlab.net/.

[43] Kai Zhang, Noah Snavely, and Jin Sun. “Leveraging vision reconstruction pipelines for satellite imagery”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. 2019, pp. 0–0.

[44] Yifeng Zhou, Henry Leung, and Martin Blanchette. “Sensor alignment with earth-centered earth-fixed (ECEF) coordinate system”. In: *IEEE Transactions on Aerospace and Electronic systems* 35.2 (1999), pp. 410–418.